(1390 REV. 5-93) US DEPT. OF COMMERCE PATENT & TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER
105045

U.S. APPLICATION NO.
(if known, sec 37 C.F.R.1.5)
09/446024

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/FR98/01442	July 6, 1998	July 7, 1997

TITLE OF INVENTION
Endogenetic Retroviral Sequences, Associated with Autoimmune Diseases or with Pregnancy Disorders

APPLICANT(S) FOR DO/EO/US
Frederic BESEME, Jean-Luc BLOND, Olivier BOUTON, Bernard MANDRAND, Francois MALLET, Herve PERRON

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.

2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.

3. ☒ This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).

4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.

5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
   b. ☒ has been transmitted by the International Bureau.
   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US)

6. ☒ A translation of the International Application into English (35 U.S.C. 371(c)(2)).

7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
   b. ☐ have been transmitted by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☐ have not been made and will not be made.

8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).

9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).

10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).

Items 11. to 16. below concern other document(s) or information included:

11. ☒ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.

12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.

13. ☒ A FIRST preliminary amendment.

    ☐ A SECOND or SUBSEQUENT preliminary amendment.

14. ☐ A substitute specification.

15. ☐ A small entity statement.

16. ☒ Other items or information: Sequence Listing.

420 Rec'd PCT/PTO 16 DEC 1999

U.S. APPLICATION NO. (if known, see 37 C.F.R. 1.5) 09/446024	INTERNATIONAL APPLICATION NO. PCT/FR98/01442	ATTORNEY'S DOCKET NUMBER 105045

17. ☒ The following fees are submitted:	CALCULATIONS	PTO USE ONLY
Basic National fee (37 CFR 1.492(a)(1)-(5)): Search Report has been prepared by the EPO or JPO.....$840.00 International preliminary examination fee paid to USPTO (37 CFR1.482).....$670.00 No international preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2)).....$690.00 Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....$970.00 International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4).....$ 96.00	$840.00	
ENTER APPROPRIATE BASIC FEE AMOUNT =	$840.00	
Surcharge of $130.00 for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

Claims	Number Filed	Number Extra	Rate		
Total Claims	20 - 20 =		X $ 18.00	$	
Independent Claims	3 - 3 =		X $ 78.00	$	
Multiple dependent claim(s)(if applicable)			+ $260.00	$	
TOTAL OF ABOVE CALCULATIONS =				$840.00	
Reduction by 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28). -				$	
SUBTOTAL =				$	
Processing fee of $130.00 for furnishing the English translation later than ☐ 20 ☐ 30 month from the earliest claimed priority date (37 CFR 1.492(f)). +				$	
TOTAL NATIONAL FEE =				$840.00	
				Amount to be refunded	$
				Charged	$

a. ☒ Check No. 105087 in the amount of $840.00 to cover the above fees is enclosed.
b. ☐ Please charge my Deposit Account No. _____ in the amount of $_____ to cover the above fees. A duplicate copy of this sheet is enclosed.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment, to Deposit Account No. 15-0461. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:
OLIFF & BERRIDGE, PLC
P.O. Box 19928
Alexandria, Virginia 22320

William Berridge

NAME: William P. Berridge
REGISTRATION NUMBER: 30,024

(1390 Rev.8-93)

420 Rec'd PCT/PTO 16 DEC 1999

U.S. APPLICATION NO. (if known, see 37 C.F.R. 1.5) 09/446024	INTERNATIONAL APPLICATION NO. PCT/FR98/01442	ATTORNEY'S DOCKET NUMBER 105045

17. ☒ The following fees are submitted:	CALCULATIONS	PTO USE ONLY
Basic National fee (37 CFR 1.492(a)(1)-(5)): Search Report has been prepared by the EPO or JPO.....$840.00 International preliminary examination fee paid to USPTO (37 CFR1.482).....$670.00 No international preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2)).....$690.00 Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....$970.00 International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4).....$ 96.00	$840.00	
ENTER APPROPRIATE BASIC FEE AMOUNT =	$840.00	
Surcharge of $130.00 for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

Claims	Number Filed	Number Extra	Rate		
Total Claims	20 - 20 =		X $ 18.00	$	
Independent Claims	3 - 3 =		X $ 78.00	$	
Multiple dependent claim(s)(if applicable)			+ $260.00	$	
TOTAL OF ABOVE CALCULATIONS =				$840.00	
Reduction by 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28). -				$	
SUBTOTAL =				$	
Processing fee of $130.00 for furnishing the English translation later than ☐ 20 ☐ 30 month from the earliest claimed priority date (37 CFR 1.492(f)). +				$	
TOTAL NATIONAL FEE =				$840.00	
				Amount to be refunded	$
				Charged	$

a. ☒ Check No. 105087 in the amount of $840.00 to cover the above fees is enclosed.
b. ☐ Please charge my Deposit Account No. _____ in the amount of $_____ to cover the above fees. A duplicate copy of this sheet is enclosed.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment, to Deposit Account No. 15-0461. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:
OLIFF & BERRIDGE, PLC
P.O. Box 19928
Alexandria, Virginia 22320

NAME: William P. Berridge
REGISTRATION NUMBER: 30,024

(1390 Rev.8-93)

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of

Frederic BESEME, Jean-Luc BLOND, Olivier BOUTON, Bernard MANDRAND, Francois MALLET, Herve PERRON

Application No.: New PCT-U.S. National Stage of PCT/FR98/01442

Filed: December 16, 1999 Docket No.: 105045

For: ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE DISEASES OR WITH PREGNANCY DISORDERS

PRELIMINARY AMENDMENT

Assistant Commissioner of Patents
Washington, D. C. 20231

Sir:

Prior to initial examination, please amend the above-identified application as follows:

IN THE TITLE:

Line 1, change "ENDOGENOUS" to --ENDOGENETIC--; and

line 2, change "AND/OR" to --OR--.

IN THE CLAIMS:

Claim 3, line 2, change "either of claims 1 and 2," to --claim 1,--.

Claim 5, lines 1-2, change "either of claims 1 and 4," to --claim 1,--.

Claim 6, lines 1-2, change "either of claims 1 and 4," to --claim 1,--.

Claim 7, lines 5-6, change "any one of claims 1 to 6" to --claim 1,--.

Claim 8, lines 4-6, change "any one of claims 1 to 6, or a nucleic fragment according to claim 7." to --claim 1.--.

Claim 10, lines 5-6, change "any one of claims 1 to 6, or a nucleic fragment according to claim 7." to --claim 1.--.

Claim 15, lines 1-3, change "claims 1 to 6, or of a nucleotide fragment according to claim 7, or of a peptide according to claim 13 or 14," to --claim 1,--.

Claim 16, lines 1-2, change "claims 1 to 6, or of a nucleotide fragment according to claim 7," to --claim 1,--.

Claim 17, lines 1-2, change "claims 1 to 6, or of a nucleotide fragment according to claim 7," to --claim 1,--.

Claim 20, lines 2-4, change "claims 1 to 6, or a nucleotide fragment according to claim 7, or a peptide according to claim 13 or 14." to --claim 1.--.

REMARKS

Claims 1-20 are pending. This Preliminary Amendment corrects typographical errors in the title and eliminates multiple dependent claims. Prompt and favorable examination is respectfully requested.

Respectfully submitted,

William P. Berridge
Registration No. 30,024

WPB:cas

OLIFF & BERRIDGE, PLC
P.O. Box 19928
Alexandria, Virginia 22320
Telephone: (703) 836-6400

ENDOGENOUS RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE DISEASES AND/OR WITH PREGNANCY DISORDERS

The present invention relates to a new nucleic material of the endogenous retroviral genomic type, various nucleotide fragments comprising it or which are obtained from said material, as well as their use as marker for at least one autoimmune disease or a pathology which is associated with it, a pathological pregnancy or an unsuccessful pregnancy.

The screening of the cDNA library with the aid of the Ppol-MSRV probe (SEQ ID NO: 29) has made it possible to detect overlapping clones allowing the reconstruction of a putative genomic RNA of 7582 nucleotides. - Reconstructed sequence is understood to mean the sequence deduced from the alignment of the overlapping clones -. This genomic RNA has the structure R-U5-gag-pol-env-U3-R. A "blastn" interrogation on several databases, with the aid of the reconstructed genome, shows that a large quantity of related genomic sequences (DNA) exist in the human genome. About 400 sequences have been identified in GenBank (cf Figure 3) and more than 200 sequences in the EST (Expressed Sequence Tag) library, the majority as antisense. These sequences are found on several chromosomes, in particular chromosomes 5, 7, 14, 16, 21, 22, X, with a high apparent concentration of LTR on the X chromosome.

The reconstructed sequence (mRNA) is integrally contained inside the genomic clone RG083M05 (gb AC00064) (9.6 kb), and exhibits 96% similarity with two discontinuous regions of this clone which also contains repeat regions at each end. The alignment of the experimental sequences corresponding to the 5' and 3' regions of the reconstructed genomic RNA with the DNA of the RG083M05 clone has made it possible to deduce an LTR sequence and to identify elements characteristic of retroviruses, in particular those

involved in reverse transcription, namely the PBS (Primer Binding Site) downstream of the 5' LTR and the PPT (PolyPurine Tract) upstream of the 3' LTR. It is observed that the U3 element is extremely short in comparison with the mammalian type C retroviruses, and comparable in size to the U3 region generally described in the type D retroviruses and the avian retroviruses. The PBS region is homologous to the PBS of the avian retroviruses, suggesting the use of the $tRNA^{Trp}$ as primer for the reverse transcription. Consequently, this new family of HERV is called HERV-W (Human Endogenous RetroVirus).

Phylogenetic analysis in the pol region has shown that the HERV-W family is phylogenetically linked to the ERV-9 and RTVL-H families, and therefore belongs to the family of type I endogenous retroviruses. Phylogenetic analysis of the open reading frame (ORF) of env shows that it is closer to the type D simian retroviruses and the avian reticuloendotheliosis retroviruses than type C mammalian retroviruses, suggesting a C/D chimeric genome structure.

The phylogenetic trees, supported by high "bootstrap" values show that the ERV-9 and HERV-W families are derived from two waves of independent insertions. Thus, the active element(s) at the origin of the HERV-W family is (are) different from that (those) from which the ERV-9 family is derived. Furthermore, the PBS of HERV-W probably uses a $tRNA^{Trp}$ whereas ERV-9 probably uses a $tRNA^{Arg}$.

Finally, the members of the HERV-W family are expressed in the placenta, whereas the ERV-9 RNAs are not detected in this tissue.

BIOLOGICAL FUNCTIONS OF HERV-W

The expression of HERV-W restricted to the placenta and the long reading frame potentially encoding a retroviral envelope make it possible to propose physiological biological functions whose impairment could be associated with pathologies.

The expression restricted to the placenta suggests that the expression of retroviral and/or nonretroviral genes under the control of the LTRs may be hormone-dependent. These genes may be adjacent, or under the control of isolated LTRs. A pathology may then result from an aberrant expression following the reactivation of a silent LTR by various factors: viral infection (for example by a member of the Herpesvirus family) or local immune activation. A polymorphism at the level of the LTRs could also promote these events.

The envelope of HERV-W could play a fusogenic role, in particular at the level of cellular subtypes of the placenta. An immunosuppressive peptide of this envelope could protect the fetus against attack by the maternal immune system. Finally, by a mechanism of saturation of receptors, the envelope of HERV-W could play a protective role against exogenous retroviral infections. The impairment of local cellular immunity may result from an immunostimulatory signal carried by the envelope. This effect may be linked to a region carrying a superantigen activity, or to the immunosuppressive region which would become immunostimulatory following either a polymorphism or a dose-effect (overexpression).

Verification of these implications and understanding of the consequences linked to an impairment of the biological functions of the endogenous LTRs or the retroviral envelope may lead to the establishment of methods of diagnosis or of monitoring:

- of states of pathological pregnancy or of unsuccessful pregnancy,
- of autoimmune diseases such as multiple sclerosis or rheumatoid arthritis.

In accordance with the present invention, there has been discovered, in the endogenous state, a new nucleic material, stated explicitly and described below, having the organization of a retrovirus, and capable of being correlated with an autoimmune disease,

or a pathology which is associated with it, with a pathological pregnancy or an unsuccessful pregnancy.

The nucleic material according to the present invention, in mRNA form, represents about 8 Kb; it is represented in Figure 1 and is described by SEQ ID NO: 11, and is represented in Figure 2 in the form of genomic DNA.

The expression "of retroviral type" is understood to mean the characteristic according to which the nucleic material considered comprises one or more nucleotide sequences related to the organization of a retrovirus, and/or to its functional or coding sequences.

This reference nucleic material is related to a human endogenous retrovirus, designated by the expression HERV-W. Consequently, it may be obtained by any appropriate technique for screening any library of human DNA, or of placental cDNA, as shown below, in particular with nucleic primers or probes synthesized so as to hybridize with all or part of SEQ ID NO: 11.

The present invention also relates to any nucleic or peptide product, obtained or derived from the reference nucleic material, according to SEQ ID NO: 11.

And finally, the invention relates to the various correlations which may be made between the abovementioned nucleic material, and/or its derived products, with any autoimmune disease and/or a pathology which is associated with it, as well as with cases of pathological pregnancy or of unsuccessful pregnancy.

"Autoimmune" is understood to mean in particular:

- multiple sclerosis
- rheumatoid arthritis
- disseminated lupus erythematosus
- insulin-dependent diabetes
- and/or pathologies which are associated with them.

The present invention relates, first of all, to a nucleic material of the retroviral genomic type, in isolated or purified state, at least partially functional or nonfunctional.

This material is characterized in that its genome comprises a reference nucleotide sequence chosen from the group including the sequences SEQ ID NOs: 1 to 15, their complementary sequences, and their equivalent sequences, in particular the nucleotide sequences exhibiting, for any sequence of 100 contiguous monomers, at least 50% and preferably at least 70%, for example at least 90% homology with respectively said sequences SEQ ID NOs: 1 to 15.

This material is also characterized in that its genome comprises a reference nucleotide sequence, encoding any polypeptide exhibiting, for any contiguous sequence of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence capable of being encoded by at least a functional part of the reference nucleotide sequence as defined above.

In particular, this material comprises a nucleic fragment inserted between two sequences corresponding respectively to the LTR region and to the gag gene for the retroviral genomic structure, in particular a nucleic fragment consisting of or comprising the sequence SEQ ID NO: 12.

The invention also relates to a nucleic material of the subgenomic retroviral type, consisting of a nucleotide sequence identical to SEQ ID NO: 11, with a deletion as exemplified by the clones cl.PH74 (SEQ ID NO: 7), cl.PH7 (SEQ ID NO: 8) and cl.Pi5T (SEQ ID NO: 9), this deletion resulting or otherwise from a splicing strategy.

The above-defined nucleic material comprises at least one functional nucleotide sequence encoding at least one retroviral protein, and/or at least one regulatory nucleotide sequence.

Next, the invention relates to any nucleotide fragment of at least 100 bases, comprising a nucleotide sequence chosen from the group comprising:

a) all the nucleotide sequences, partial and complete, of a nucleic material as defined above

b) all the nucleotide sequences, partial and complete, of a clone chosen from the group including the clones:

- cl.6A2 (SEQ ID NO: 1)
- cl.6A1 (SEQ ID NO: 2)
- cl.7A16 (SEQ ID NO: 3)
- cl.Pi22 (SEQ ID NO: 4)
- cl.24.4 (SEQ ID NO: 5)
- cl.C4C5 (SEQ ID NO: 6)
- cl.PH74 (SEQ ID NO: 7)
- cl.PH7 (SEQ ID NO: 8)
- cl.Pi5T (SEQ ID NO: 9)
- cl.44.4 (SEQ ID NO: 10)
- HERV-W (SEQ ID NO: 11)
- cl.6A5 (SEQ ID NO: 12)
- cl.7A20 (SEQ ID NO: 13)
- cl.7A21 (SEQ ID NO: 14)
- LTR (SEQ ID NO: 15)

c) the sequences which are respectively complementary to the sequences according to a) and b)

d) the sequences which are respectively equivalent to the sequences according to a) to c), in particular the nucleotide sequences exhibiting, for any sequence of 100 contiguous monomers, at least 50%, and preferably at least 70%, or even better at least 80%, for example at least 90% homology with the sequences a) to c).

The invention also relates to any nucleic probe for the detection of a nucleic material, inserted or otherwise into a nucleic acid, characterized in that it is capable of hybridizing specifically with a nucleic material, as defined above.

Such a probe comprises a marker or otherwise.

The invention also relates to a nucleic primer for the amplification by polymerization of an RNA or of a DNA, characterized in that it comprises a nucleotide sequence capable of hybridizing specifically with a nucleic material or a nucleic fragment, as defined above.

By way of example, a nucleic probe or nucleic primer according to the invention is characterized in that it consists of a nucleotide sequence chosen from the group including SEQ ID NOs: 16 to 28.

The invention also relates to any RNA or DNA, and in particular a replication vector, comprising a nucleotide fragment, as defined above.

The invention also relates to any peptide encoded by any open reading frame belonging to a nucleotide fragment, as defined above, in particular polypeptide, for example oligopeptide forming an antigenic determinant recognized by sera from patients affected by an autoimmune disease, or a pathology which is associated with it, or from patients having a pathological pregnancy or an unsuccessful pregnancy.

By way of example, this polypeptide is encoded by a nucleotide fragment comprising an open reading frame encoding one or more retroviral ENV proteins.

Finally, the invention relates to:

- the use of a nucleic material, or of a nucleotide fragment, or of a peptide defined above, as previously defined, as molecular marker for an autoimmune disease or for a pathology which is associated with it, for pathological pregnancy or unsuccessful pregnancy;

- the use of a nucleic material, or of a nucleotide fragment, as defined above, as chromosomal marker for susceptibility to an autoimmune disease or for a pathology which is associated with it, or for a risk of a pathological pregnancy or of an unsuccessful pregnancy;

- the use of a nucleic material, or of a nucleotide fragment, as defined above, as proximity

marker for a gene for susceptibility to an autoimmune disease or to a pathology which is associated with it, or to a risk of a pathological pregnancy or of an unsuccessful pregnancy.

The invention also relates to a method for the molecular labeling of an autoimmune disease or of a pathology which is associated with it, of pathological pregnancy or of unsuccessful pregnancy, characterized in that any nucleotide fragment, as defined above, either in RNA form or in DNA form, is identified and/or quantified in any biological body material, in particular body fluid.

By way of example, according to such a method, cells expressing a nucleotide fragment, as defined above, are detected in said biological body material.

The invention relates to a diagnostic and/or therapeutic application of a nucleic material, of a nucleotide fragment or of a peptide defined above, and as such, another subject of the invention is a diagnostic composition or a therapeutic composition comprising said material, said fragment or said peptide.

Before detailing the invention, various terms used in the description and the claims are now defined:

- human virus is understood to mean a virus capable of infecting or of being harbored by a human being,

- taking into account all the natural or induced variations and/or recombinations which may be encountered in the practical implementation of the present invention, the subjects thereof, defined above and in the claims, have been expressed comprising the equivalents or derivatives of the different biological materials defined below, in particular the homologous nucleotide or peptide sequences,

- the variant of a virus or of a pathogenic and/or infective agent according to the invention comprises at least one antigen recognized by at least one antibody directed against at least one corres-

ponding antigen of said virus and/or of said pathogenic and/or infective agent, and/or a genome of which any part is detected by at least one hybridization probe, and/or at least one nucleotide amplification primer specific for said virus and/or pathogenic and/or infective agent, in particular a genome belonging to the HERV-W family, under determined hybridization conditions well known to persons skilled in the art,

- according to the invention, a nucleotide fragment or an oligonucleotide or a polynucleotide is a stretch of monomers, or a biopolymer, characterized by the sequence, informational or otherwise, of the natural nucleic acids, capable of hybridizing with any other nucleotide fragment under predetermined conditions, it being possible for the stretch to contain monomers of different chemical structures and to be obtained from a natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis; a nucleotide fragment may be identical to a genomic fragment of an element of the HERV-W family considered by the present invention, in particular a gene for the latter, for example pol or env in the case of said element;

- thus, a monomer may be a natural nucleotide of a nucleic acid, whose constituent elements are a sugar, a phosphate group and a nitrogen base; in RNA, the sugar is ribose, in DNA, the sugar is 2-deoxyribose; depending on whether DNA or RNA is involved, the nitrogen base is chosen from adenine, guanine, uracil, cytosine, thymine; or the nucleotide may be modified in at least one of the three constituent elements; by way of example, the modification may take place at the level of the bases, generating modified bases such as inosine, 5-methyldeoxycytidine, deoxyuridine, 5-(dimethylamino)deoxyuridine, 2,6-diaminopurine, 5-bromodeoxyuridine and any other modified base promoting hybridization; at the level of the sugar, the modification may consist in the replacement of at least one deoxyribose with a

polyamide, and at the level of the phosphate group, the modification may consist in its replacement with esters, in particular chosen from diphosphate, alkyl and arylphosphonate and phosphorothioate esters,

- "functional" is understood to mean the characteristic according to which a nucleotide sequence, a nucleic material or a nucleotide fragment comprises an "an informational sequence",

- "informational sequence" is understood to mean any ordered sequence of monomers whose chemical nature and the order in a reference direction, constitute or otherwise a functional information of the same quality as that of the natural nucleic acids, for example a reading frame encoding a protein, a regulatory sequence, a splicing site or a recombination site,

- hybridization is understood to mean the process during which, under appropriate operating, in particular, stringency, conditions, two nucleotide fragments, having sufficiently complementary sequences, pair to form a complex, in particular double or triple, structure, preferably in the form of a helix,

- a probe comprises a nucleotide fragment synthesized in particular by the chemical or polymerization route, or obtained by enzymatic digestion or cleavage of a longer nucleotide fragment, comprising at least six monomers, advantageously from 10 to 100 monomers, preferably 10 to 30 monomers, and possessing a hybridization specificity under determined conditions; preferably, a probe possessing less than 10 monomers is not used alone, but is used in the presence of other probes equally short in size or otherwise; under certain specific conditions, it may be useful to use probes larger than 100 monomers in size; a probe may in particular be used for diagnostic purposes and it will include for example capture and/or detection probes,

- the capture probe may be immobilized on a solid support by any appropriate means, that is to say

directly or indirectly, for example by covalence or by passive adsorption,

- the detection probe may be labeled by means of a marker chosen in particular from radioactive isotopes, enzymes particularly chosen from peroxidase and alkaline phosphatase and those capable of hydrolyzing a chromogenic, fluorigenic or luminescent substrate, chromophoric chemical compounds, chromogenic, fluorigenic or luminescent compounds, nucleotide base analogs, and biotin,

- the probes used for diagnostic purposes of the invention may be used in all the hybridization techniques known to persons skilled in the art, and in particular the techniques termed "DOT-BLOT", "SOUTHERN BLOT", "NORTHERN BLOT" which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as target, the SANDWICH technique; advantageously, the SANDWICH technique is used in the present invention, comprising a specific capture probe and/or a specific detection probe, it being understood that the capture probe and the detection probe must have a nucleotide sequence which is at least partially different,

- any probe according to the present invention may hybridize in vivo or in vitro with RNA and/or with DNA, to block the phenomena of replication, in particular translation and/or transcription, and/or to degrade said DNA and/or RNA,

- a primer is a probe comprising at least six monomers, and advantageously from 10 to 30 monomers, possessing a hybridization specificity under determined conditions, for the initiation of an enzymatic polymerization, for example in an amplification technique such as PCR (Polymerase Chain Reaction), in an extension method such as sequencing, in a reverse transcription method and the like,

- two nucleotide or peptide sequences are said to be equivalent or derived from each other, or relative to a reference sequence, if functionally the corresponding biopolymers may play substantially the

same role, without being identical, in relation to the application or use considered, or in the technique in which they are used; in particular equivalent are two sequences obtained because of the natural variability within the same individual, or the natural diversity from one individual to another within the same species, in particular spontaneous mutation of the species from which they were identified, or induced mutation, as well as two homologous sequences, the homology being defined below,

- "variability" is understood to mean any modification, spontaneous or induced, of a sequence, in particular by substitution, and/or insertion, and/or deletion of nucleotides and/or of nucleotide fragments, and/or extension and/or shortening of the sequence at at least one of the ends; an unnatural variability may result from the genetic engineering techniques used, for example from the choice of the synthetic primers, degenerate or otherwise, selected for amplifying a nucleic acid; this variability may result in modifications of any starting sequence, considered as reference, and which may be expressed by a degree of homology relative to said reference sequence,

- homology characterizes the degree of identity of two nucleotide or peptide fragments compared; it is measured by the percentage identity which is in particular determined by direct comparison of nucleotide or peptide sequences, relative to reference nucleotide or peptide sequences,

- this percentage identity was specifically determined for the nucleotide fragments, in particular clones within the present invention, and obtained from the same individual; by way of nonlimiting example, the lowest percentage identity observed between the different clones from the same individual (cf SEQ ID NOs: 13 and 14) is at least 90% and the lowest percentage identity observed between the different clones of two individuals is at least 80%,

- any nucleotide fragment is said to be equivalent to or derived from a reference fragment if it exhibits a nucleotide sequence equivalent to the sequence of the reference fragment; according to the above definition, particularly equivalent to a reference nucleotide fragment are:

(a) any fragment capable of at least partially hybridizing with the complement of the reference fragment,

(b) any fragment whose alignment with the reference fragment leads to identical contiguous bases being identified in a larger number than with any other fragment obtained from another taxonomic group,

(c) any fragment resulting or capable of resulting from the natural variability within the same individual, and from the natural diversity from one individual to another within the same species, from which it is obtained,

(d) any fragment capable of resulting from genetic engineering techniques applied to the reference fragment,

(e) any fragment, containing at least eight contiguous nucleotides, encoding a peptide homologous or identical to the peptide encoded by the reference fragment,

(f) any fragment different from the reference fragment by insertion, deletion, substitution of at least one monomer, extension, or shortening at at least one of its ends; for example, any fragment corresponding to the reference fragment, flanked at at least one of its ends by a nucleotide sequence not encoding a polypeptide,

- partial or complete nucleotide sequence of a reference nucleic material is also understood to mean any sequence associated by co-encapsidation, or by coexpression, or recombined with said reference nucleic material,

- polypeptide is understood to mean in particular any peptide of at least two amino acids, in

particular oligopeptide or a protein, extracted, separated or substantially isolated or synthesized, through the intervention of human hands, in particular those obtained by chemical synthesis, or by expression in a recombinant organism,

- polypeptide partially encoded by a nucleotide fragment is understood to mean a polypeptide having at least three amino acids encoded by at least nine contiguous monomers contained in said nucleotide fragment,

- an amino acid is said to be analogous to another amino acid when their respective physico-chemical characteristics, such as polarity, hydrophobicity and/or basicity, and/or acidity, and/or neutrality, are substantially the same; thus, a leucine is analogous to an isoleucine,

- any polypeptide is said to be equivalent to or derived from a reference polypeptide if the compared polypeptides have substantially the same properties, and in particular the same antigenic, immunological, enzymological and/or molecular recognition properties; particularly equivalent to a reference polypeptide is:

(a) any polypeptide possessing a sequence in which at least one amino acid has been substituted with an analogous amino acid;

(b) any polypeptide having an equivalent peptide sequence obtained by natural or induced variation of said reference polypeptide, and/or of the nucleotide fragment encoding said polypeptide,

(c) a mimotope of said reference polypeptide,

(d) any polypeptide in whose sequence one or more amino acids of the L series are replaced by an amino acid of the D series, and vice versa,

(e) any polypeptide into whose sequence a modification of the side chains of the amino acids has been introduced, such as for example an acetylation of the amine functions, a carboxylation of the thiol functions, an esterification of the carboxyl functions,

(f) any polypeptide in whose sequence one or more peptide bonds have been modified, such as for example the carba, retro, inverse, retro-inverse, reduced and methyleneoxy bonds,

(g) any polypeptide of which at least one antigen is recognized by an antibody directed against a reference polypeptide,

- the percentage identity characterizing the homology between two compared peptide fragments is, according to the present invention, at least 80% and preferably at least 90%.

The expressions relating to order which are used in the present description and the claims, such as "first nucleotide sequence" are not selected to express a particular order, but to define the invention more clearly.

Detection of a substance or agent is understood to mean hereinafter both an identification and a quantification, or a separation or isolation of said substance or of said agent.

The invention will be understood more clearly upon reading the detailed description which follows, made with reference to the appended figures in which:

- Figure 1 represents, on the one hand, the organization of the endogenous retroviral material discovered according to the present invention, in the form of a putative genomic mRNA, and, on the other hand, the location of the clones used according to the present invention, relative to this organization; the scales for length are expressed in Kb; the flanking regions (5' UTR and 3' UTR) are indicated in hatched boxes; the regions repeated in these two flanking regions are indicated by black arrows; the regions corresponding to the gag, pol and env genes are indicated in black, white and gray respectively; the position of the Ppol-MSRV probe is indicated;

- Figure 2 represents a possibility of genetic organization (DNA), illustrated by the clone RG083M05, and a splicing strategy linking to this sequence, the

experimental clones (mRNA); this figure also shows the splicing sites observed with reference to the retroviral organization; additionally indicated in this figure are:

the location of the probes used (Pgag-LB19, Ppro-E, Ppol-MSRV and Penv-C15);

the splice donor sites (DS1 and DS2) and acceptor sites (AS1 to AS3);

the sequences obtained from the clone RG083M05, in the lower-case boxes, and the sequences derived from experimental placental clones (mRNA), in the upper-case boxes;

the putative ORFs (ORF1, ORF2 and ORF3); and

an insert of 2 Kb present in DNA form but not detected in RNA form, represented in the form of vertical hatches.

The other conventions used in this figure are the same as those for Figure 1.

- Figure 3 gives a representation of genomic (DNA) clones corresponding to the isolated cDNA clones; indicated in this figure are:

the percentage similarity with respect to the reconstructed genomic RNA (Recons RNA);

the presence of repeat sequences at each end of these genomes (repeats); and

the presence and the size of the open reading frames (ORFs).

- Figure 4 represents a phylogenetic analysis identifying the HERV-W family.

- Figure 5 represents the alignment of the 5' and 3' flanking regions of the clone RG083M05 with the terminal 5' and/or 3' regions of some placental clones; the CAAC tandem flanking the 3' and 5' LTRs is doubly underlined under the DNA sequences, the consensus LTR sequence of 783 bp (base pairs) is indicated under the alignment; the PPT upstream of the 5' end of LTR and the PBS downstream of the 3' end of LTR are indicated; the U3R and U5 regions are indicated; the sites corresponding to the binding of the transcription

factor are underlined and numbered from 1 to 6; the region -73 to 284 corresponds to the sequence evaluated in "CAT assay"; * corresponds to putative sites for "capping"; [polyA] indicates the polyadenylation signal.

- Figure 6 represents a putative sequence of a HERV-W envelope polypeptide (ORF1) obtained from 3 different placental cDNA clones; the leader peptide (L), the surface protein (SU) and the transmembrane protein (TM) are indicated by arrows; the hydrophobic fusion peptide and the transmembrane carboxy region are underlined by a single line and a double line, respectively; the immunosuppression region is indicated in italics; the potential glycosylation sites are indicated by dots; the divergent amino acids are indicated on the bottom line; Figure 6 also presents the open reading frames corresponding to ORF2 and ORF3 as described in Figure 2, and more particularly their homologies with the retroviral regulatory genes.

The nucleic material previously presented explicitly was discovered and characterized at the end of the experimental protocol described below, it being understood that this protocol cannot limit the scope of the present invention and of the accompanying claims.

Example 1

Isolation and sequencing of overlapping cDNA fragments

The information relating to the organization of HERV-W were obtained by testing a placental cDNA library (Clontech cat#HL5014a) with the probes Ppol-MSRV (SEQ ID NO: 29) and Penv-C15 (SEQ ID NO: 31) (cf Example 8), and then performing a "gene walking" technique with the aid of the new sequences obtained. The experiments were carried out with reference to the recommendations of the supplier of the library. PCR amplifications on DNA were also exploited in order to understand this organization.

A number of clones were selected and sequenced, cf Figure 1:

- clone cl.6A2 (SEQ ID NO: 1): untranslated 5' region of HERV-W and part of gag

- clone cl.6A1 (SEQ ID NO: 2): gag and part of pol

- clone cl.7A16 (SEQ ID NO: 3): 3' region of pol

- clone cl.Pi22 (SEQ ID NO: 4): 3' region of pol and beginning of env

- clone cl.24.4 (SEQ ID NO: 5): spliced RNA comprising part of the untranslated 5' region of HERV-W, the end of pol and the 5' region of env

- clone cl.C4C5. (SEQ ID NO: 6): end of env and untranslated 3' region of HERV-W

- clone cl.PH74 (SEQ ID NO: 7): subgenomic RNA: untranslated 5' region of HERV-W, end of pol, env and untranslated 3' region of HERV-W

- clone cl.PH7 (SEQ ID NO: 8): multispliced RNA: untranslated 5' region of HERV-W, end of env and untranslated 3' region of HERV-W.

- clone cl.Pi5T (SEQ ID NO: 9): partial pol gene and U3-R region

- clone cl.44.4 (SEQ ID NO: 10): R-U5 region, gag gene and partial pol gene.

With the aid of these clones, by carrying out sequence alignments, a model of complete sequence of HERV-W was produced. The spliced RNAs were identified as well as the potential splice donor and acceptor sites. This set of information is shown in Figure 2. Through a study of similarity with existing retroviruses, the LTR, gag, pol and env entities were defined.

The putative genetic organization of HERV-W in RNA form is the following (SEQ ID NO: 11):

gene 1..7582

location of the clones on the reconstructed genomic RNA sequence

cl.6A2 (1321 bp) 1-1325;

cl.PH74 (535+2229= 2764 bp) 72-606 and 5353-7582;

cl.24.4 (491+1457= 1948 bp); 115-606 and 5353-6810;
cl.44.4 (2372 bp) 115-2496;
cl.PH7 (369+297= 666 bp) 237-606 and 7017-7313;
cl.6A1 (2938 bp) 586-3559.;
cl.Pi5T (2785+566= 3351 bp) 2747-5557 and 7017-7582;
cl.7A16 (1422 bp) 2908-4337;
cl.Pi22 (317+1689= 2006 bp) 3957-4273 and 4476-6168;
cl.C4C5 (1116 bp) 6467-7582

5'LTR 1..120
/note="R of 5'LTR (5' end uncertain"

121..575
/note="U5 of 5'LTR"

various 579..596
/note="PBS primer binding site for tRNA-W"

various 606
/note="splice junction (splice donor site ATCCAAAGTG-GTGAGTAATA and splice acceptor site CTTTTTTCAG-ATGGGAAACG clone RG083M05, GenBank accession AC000064)"

various 5353
/note="splice acceptor site for ORF1 (env)"

various 5560
/note="splice donor site"

ORF 5581..7194
/note="ORF1 env 538 AA"
/product-="envelope"

various 7017
/note="splice acceptor site for ORF2 and ORF3"

ORF 7039..7194
/note="ORF2 52 AA"

ORF 7112..7255
/note="ORF3 48 AA"

various 7244..7254
/note="PPT polypurine tract"

3'LTR 7256..7582
/note-="U3-R of 3' LTR (U3-R junction indeterminate)
various 7563..7569
polyadenylation signal

Example 2:

Identification of genomic (DNA) clones corresponding to the isolated DNA clones

A "blastn" interrogation of several databases, with the aid of the reconstructed genome, shows that a large quantity of related sequences exist in the human genome. About 400 sequences were identified in GenBank and more than 200 sequences in the EST library, and the majority as antisense. The 4 sequences most significant in size and in similarity, illustrated in Figure 3, are the following genomic (DNA) clones:

the human clone RG083M05 (gb AC000064) whose chromosomal location is 7q21-7q22,

the human clone BAC378 (gb U85196, gb AE000660) corresponding to the alpha delta locus of the T cell receptor, located in 14q11-12,

the human cosmid Q11M15 (gb AF045450) corresponding to the 21q22.3 region of chromosome 21,

the cosmid U134E6 (embl Z83850) on chromosome Xq22.

The location of the aligned regions for each of the clones is indicated and the affiliation to a chromosome is indicated in square brackets. The percentage similarity (without broad deletions) between the 4 sequences and the reconstructed genomic RNA is indicated, as well as the presence of repeat sequences at each end of the genome and the size of the largest reading frames (ORF). Repeat sequences are found at the ends of 3 of these clones. The reconstructed sequence is integrally contained inside the clone RG083M05 (9.6 Kb) and exhibits a 96% similarity. However, the clone RG083M05 exhibits an insert of 2 Kb situated immediately downstream of the untranslated 5' region (5' UTR). This insert is also found in two other

genomic clones which exhibit a deletion of 2.3 Kb immediately upstream of the untranslated 3' region (3' UTR). No clone contains the three functional reading frames (ORFs) gag, pol and env. The clone RG083M05 shows an ORF of 538 amino acids (AA) corresponding to a whole envelope. The cosmid Q11M15 contains two large contiguous ORFs of 413 AA (frame 0) and 305 AA (frame +1) corresponding to a truncated pol polyprotein.

Example 3

Phylogenetic analysis

A phylogenetic analysis was carried out at the level of the nucleic acids on 11 different subregions of the reconstructed genomic RNA, and at the protein level on 2 different subregions of env. All the trees obtained exhibit the same topology regardless of the region studied. This is illustrated in Figure 4 at the level of the nucleic acids in the most conserved LTR and pol regions between the sequences obtained and ERV-9 and RTLV-H. The trees clearly show that the experimental sequences describe a new family distinct from ERV-9 and very distinct from RTLV-H as underlined by the "bootstrap" analysis. These sequences are found on several chromosomes, in particular chromosomes 5, 7, 14, 16, 21, 22 and X with a high apparent concentration of LTR on the X chromosome.

Comparison at the protein level between the most conserved regions of the retroviral env proteins shows that the HERV-W family is closer to the type D simian retroviruses and the avian reticuloendotheliosis retroviruses than the type C mammalian retroviruses.

This suggests a C/D chimeric genomic structure.

Example 4

Identification of the LTR, PPT and PBS elements

The reconstructed sequence (RNA) is integrally contained inside the genomic clone RG083M05 (9.6 Kb) and exhibits a 96% similarity with two discontinuous regions of this clone which also contains repeat regions at each end. The alignment of the experimental

sequences corresponding to the 5′ and 3′ regions of the genomic RNA reconstructed with the DNA of the clone RG083M05 [5′(5-RG-28000-28872) and 3′(3-RG-37500-38314)] made it possible to deduce an LTR sequence and to identify elements characteristic of the retroviruses, in particular those involved in the reverse transcription, namely PBS downstream of the 5′ LTR and the PPT upstream of the 3′ LTR (cf Figure 5). It is observed that the U3 element is extremely short in comparison with that observed in the mammalian type C retroviruses, and is comparable in size to the U3 region generally described in the type D retroviruses and the avian retroviruses. The region corresponding to bases 2364 to 2720 of the clone cl.PH74 (SEQ ID NO: 7) was amplified by PCR and subcloned into the vector pCAT3 (Promega) in order to carry out the evaluation of the promoter activity. A significant activity was found in HeLa cells by the so-called "CAT assay" method showing the functionality of the promoter sequence of the LTR.

The PBS region is homologous to the PBS of the avian retroviruses.

Example 5

Genetic organization and regulation of expression

Organization in DNA form

PCR amplifications were carried out on whole HERV-W clones recovered on human genomic library (see Example 1 for the mode of production), using the following oligonucleotide pairs:

U5 4992 (SEQ ID NO: 16), GAG 4619 (SEQ ID NO: 17)
GAG 4782 (SEQ ID NO: 18), POL 3167 (SEQ ID NO: 19)
POL 3390 (SEQ ID NO: 20), POL 5144 (SEQ ID NO: 21)
POL 5145 (SEQ ID NO: 22), U5 4991 (SEQ ID NO: 23).

The PCRs were carried out under the following conditions:

oligonucleotides at the concentration of 0.33 microMolar

TAQ polymerase buffer Boerhinger 1X
0.5 unit of TAQ polymerase Boerhinger
mixture of dNTP at 0.25 mM each
0.5 mg of human DNA
final volume 100 ml

PCR conditions (95°C, 5 min) × 1, (95°C, 30 sec + 54°C, 30 sec + 72°C 3 min) × 35.

The PCR products were then deposited on 1% agarose gel to be analyzed after migration. The set of PCRs gives amplification fragments of the expected size, except for the LTR-4991—gag-4619 PCR which gives a fragment of size greater by about 2 Kb relative to the expected size (deduced from cDNAs from the placental library). The reconstruction of HERV-W in endogenous DNA form therefore represents an entity of about 10 Kb.

After cloning, sequencing and analysis of the PCR-4992 gag-4619, the presence of a region of insertion is observed between LTR and gag of SEQ ID NO: 12 (clone cl.6A5). This region does not correspond to an untranslated traditional region of a retrovirus: no ψ or PBS region.

The products of PCR pol-3390, pol-5144 were also cloned and two of the clones obtained were sequenced. The result of these sequences is given by the clones cl.7A20 (SEQ ID NO: 13) and cl.7A21 (SEQ ID NO: 14). Comparison of these two nucleotide sequences gives a score of 90% homology for the relevant region, thus showing the variability of HERV-W in the same individual.

HERV-W in DNA form is proposed in Figure 2.

General organization: transcription process

The various cDNA clones having been obtained, results acquired in PCR on DNA, there is deduced:

- a DNA organization of 10 Kb possessing an insertion sequence of 2 Kb between LTR and gag.

The result of PCR on DNA showing the presence of an insert of 2 Kb between the LTR and gag regions suggests that the cDNAs isolated from the placenta are

obtained from the expression of a genome of the RG083M05 type.

- an RNA organization of 8 Kb resulting from a transcription of 10 Kb followed by a splicing between LTR and gag making it possible to restore a continuity FR (Flanking Region) 5' gag, and thus giving an RNA of 8 Kb as identified in Northern blotting.

The probes gag (Pgag-LB19, SEQ ID NO: 30) and protease (Ppro-E, SEQ ID NO: 32) reveal an RNA having a size close to 8 Kb, the probe Penv-C15 (SEQ ID NO: 31) reveals, in addition, an RNA close to 3.1 Kb. Two probes defined in the untranslated 5' region, obtained by screening of the cDNA library reported above (probe P5'-gag-cl.6A2 derived from the clone cl.6A2 and probe P5'-env-cl.24.4 derived from the clone cl.24.4) reveal the preceding two RNAs and an RNA of about 1.3 Kb. This distribution of the RNAs is typical of complex retrovirus transcripts: a genomic RNA encoding gag-pro-pol, a subgenomic RNA encoding the envelope, and one or more multispliced RNAs potentially encoding regulatory genes.

The half-life of such an RNA (LTR-R-U5-Insertion-GAG-POL-ENV-U3-R-HERV-W) is probably very short, because no RNA of 10 Kb is detected in Northern blotting. By analyzing and comparing sequences, the potential splice donor sites (DS1 and DS2) and acceptor sites were defined and described in Figure 2.

Example 6

Transcription in healthy tissues

Various healthy human tissues were tested by the Northern-blot technique (Human Multiple Tissue Northern Blot, Clontech cat# 7760-1), with the aid of the probes Ppol-MSRV (SEQ ID NO: 29), Pgag-LB19 (SEQ ID NO: 30), Penv-C15 (SEQ ID NO: 31), Ppro-E (SEQ ID NO: 32), P5'-gag-cl.6A2 and P5'-env-cl.24.4, labeled as described in Example 1. The experiments were carried out following the recommendations of the manufacturers, and the autoradiographs were exposed for 5 days. Analysis of the results reveals transcription

products only in the placenta, and in none of the other human tissues tested (heart, brain, lungs, liver, skeletal muscle, kidney and pancreas).

Using an RNA Dot-Blot technique (Clontech: Human RNA Master Blot Cat# 7770-1), and using the experimental protocol recommended by the manufacturer, about forty other tissues, including fetal tissues, were tested: only the placenta gives a specific response after hybridization with the probes Pgag-LB19 (SEQ ID NO: 30) and Penv-C15 (SEQ ID NO: 31).

It is observed that a signal is observed in the kidney in RNA Dot-Blot, which is infirmed by the Northern-blot analysis.

Example 7

Identification of an mRNA encoding an envelope and the means for detecting it specifically

The screening of a placental cDNA library with the aid of a probe defined in the untranslated 5' region made it possible to isolate a cDNA defined by an untranslated 5' region (5' NTR), a splicing junction, a coding sequence, an untranslated 3' region (3' NTR) and a polyadenylated tail, cl.PH74 (SEQ ID NO: 7). This clone corresponds to a spliced RNA encoding an envelope. By comparing sequences between this cDNA and the endogenous HERV-W model proposed according to Figure 2, a splicing junction is identified on the mRNA, a splicing junction placing in continuity the 5' NTR region and the env gene, leading to the production of a spliced subgenomic RNA encoding the envelope gene. This information made it possible to define an oligonucleotide specific for this mRNA by choosing a location situated on the splicing site (Oligo 5307, according to SEQ ID NO: 24).

The identification of this joining region makes it possible to establish a method of discriminating between endogenous retroviral RNA and DNA, using, in a PCR, an oligonucleotide defined on this joining region, in particular an oligonucleotide chosen from the env gene (Oligo 4986, according to SEQ ID NO: 25).

The PCRs were carried out under the following conditions:

oligonucleotides at the concentration of 0.33 microMolar

TAQ polymerase buffer Boerhinger 1X

0.5 unit of TAQ polymerase Boerhinger

mixture of dNTP at 0.25 mM each

0.5 mg of human DNA

final volume 100 ml

On 10 different DNAs tested, this type of PCR did not make it possible to obtain amplification products. On the other hand, on cDNA derived from placental RNA or from cells expressing HERV-W, this PCR gives an amplification product. This result therefore confirms the specifically RNA nature of this subgenomic fragment.

Example 8

Identification of coding sequences contained in a specific mRNA

The splicing strategy described in Example 5 is compatible with the presence of three reading frames ORF1 (SEQ ID NO: 33), ORF2 (SEQ ID NO: 34) and ORF3 (SEQ ID NO: 35) (cf Figure 6).

The screening of a placental cDNA library made it possible to isolate a cDNA (SEQ ID NO: 7, cl.PH74) defined by an untranslated 5' region (5' NTR), a splicing junction, a coding sequence, an untranslated 3' region (3' NTR) and a polyadenylated tail. The coding sequence is 538 amino acids (SEQ ID NO: 33). The analyses carried out on databanks make it possible to identify characteristics of a complete retroviral envelope: initiation of translation of an envelope polyprotein, of a highly hydrophobic leader peptide of about 21 amino acids, of a surface protein SU, of a transmembrane protein TM. These two protein entities exhibit different potential glycosylation sites. An immunosuppressive region is identified within the TM protein.

22 bp and 95 bp upstream of the splice acceptor site, two initiation codons were respectively found which were capable of directing the synthesis of 52 AA (ORF2, SEQ ID NO: 34) and of 48 AA (ORF3, SEQ ID NO: 35). ORF2 consists of part of the carboxy-terminal end of env and ORF3 corresponds to a different but overlapping translation.

No significant homology was found by "blast" interrogation. However, an LFASTA interrogation in a sub-databank limited to the Retroviridae, ORF2 and ORF3 showed a percentage identity of 35% with, respectively, Rex of the human and primate lymphotropic T virus, and with Tat of the simian immunodeficiency virus.

Example 9

Complexity of the HERV-W family

The number of copies present in the human genome of each of the sequences is evaluated by a Dot-Blot technique, with the aid of the probes Pgag-LB19 (SEQ ID NO: 30), Ppro-E (SEQ ID NO: 32) and Penv-C15 (SEQ ID NO: 31).

Each of the probes is denatured and deposited on a Hybond N+ membrane in an amount of 2.5, 5, 10, 25, 50, 100 pg per deposit. 0.5 mg of human DNA are also deposited on the same membrane. The membranes are dried for 2 hours under vacuum at 80°C. The membranes are then hybridized with the deposited probe. The techniques for labeling the probes, for hybridization and for washing the membranes are the same as for the Southern blotting. After autoradiography of the membranes, levels of signal intensity which are proportional to the deposits on the membrane are observed. After cutting out the hybridization zones, scintillation counting is carried out. By comparison between the dilution series for the probe deposited on the membrane and the result obtained with the human DNA, it is possible to evaluate the number of copies per haploid genome of each of the regions covered by the probes:

- the number of endogenous gag is evaluated from 56 to 112 copies (76)

- the number of endogenous protease is evaluated from 166 to 334 copies (260)

- the number of endogenous env is evaluated at less than 52 copies (13).

The screening of 10^6 clones of a human placental DNA library (Clontech cat# Hl5014b) made it possible to count 144 clones recognized by the probe Pgag-LB19, and 64 clones recognized by the probe Penv-C15. 13 clones hybridized conjointly with the probes Penv-C15 and Pgag-LB19 were isolated, confirming the presence of several copies of a genome possessing both gag and env, without consideration of functionality.

The nucleic material, the nucleotide sequences and the peptides or proteins which may be expressed by said materials and sequences may be used to detect, predict, treat and monitor any autoimmune disease, and the pathologies which are associated with it, as well as in cases of pathological pregnancy or of unsuccessful pregnancy.

Indeed, the objective and experimental data make it possible to link retrovirus and autoimmune diseases and retrovirus and pregnancy disorders:

(1) common mechanisms are used in the retroviral pathologies and in autoimmune diseases (presence of autoantibodies, of immune complexes, cellular infiltration of certain tissues, neurological disorders).

(2) pathological disorders comparable to certain autoimmune diseases appear during infections with HIV and HTLV retroviruses (Sjögren syndrome, disseminated lupus erythematosus, rheumatoid arthritis and the like).

(3) a reverse transcriptase activity was detected and retroviral-type particles were observed in the cell culture supernatants of patients suffering from multiple sclerosis (Perron et al., Res. Virol.

1989; 140: 551-561/Lancet 1991; 337: 862-863/Res. Virol. 1992; 143: 337-350) or from rheumatoid arthritis.

(4) autoimmune or chronic inflammatory animal pathologies are linked to endogenous retroviruses; some of them are used as animal models of human diseases (insulin-dependent diabetes, disseminated lupus erythematosus).

(5) significant levels of endogenous anti-retrovirus antibodies have been described in the context of autoimmune, systemic or inflammatory diseases; other data of this nature were communicated by several authors at the IVth European meeting on endogenous retroviruses (Uppsala, October 1996). According to Venables (communiqués of the IVth European meeting on endogenous retroviruses, Uppsala, October 1996), a significantly high level of anti-HERV-H antibodies are found during pregnancy but also in the context of various autoimmune disorders such as Sjögren syndrome, disseminated lupus erythematosus or rheumatoid arthritis, without, however, any proof of its direct involvement being provided up until now.

The involvement of the retroviruses in the autoimmune phenomenon remains compatible with the multifactorial character of the autoimmune, systemic or inflammatory diseases which confront genetic, hormonal, environmental and infectious factors.

The particles observed in the cell culture supernatants from patients suffering from multiple sclerosis (Perron et al., Res. Virol. 1989; 140: 551-561/Lancet 1991; 337: 862-863/Res. Virol. 1992; 143: 337-350) or from rheumatoid arthritis (unpublished data) may result from the expression: (i) of an endogenous retrovirus competent for replication, (ii) of several defective endogenous retroviruses cooperating by a phenomenon of transcomplementation or (iii) of an exogenous retrovirus.

All these observations make it possible to use and consider the above-described biological material as

marker for an autoimmune disease or for pregnancy disorders.

In particular, the following labeling techniques are considered:

- screening of the human genome with high-stringency hybridization probes derived from the nucleic material described above,
- direct amplification of genomic DNA by PCR, using primers specific for the region considered
- analysis of the flanking regions of foreign cellular genes.

CLAIMS

1. Nucleic material of the retroviral genomic type, in isolated or purified state, at least partially functional or nonfunctional, whose genome comprises a reference nucleotide sequence chosen from the group including the sequences SEQ ID NOs: 1 to 15, their complementary sequences, and their equivalent sequences, in particular the nucleotide sequences exhibiting, for any sequence of 100 contiguous monomers, at least 70% and preferably at least 90% homology with respectively said sequences SEQ ID NOs: 1 to 15.

2. Nucleic material of the retroviral genomic type, in isolated or purified state, at least partially functional or nonfunctional, whose genome comprises a reference nucleotide sequence, encoding any polypeptide exhibiting, for any contiguous sequence of at least 30 amino acids, at least 80%, and preferably at least 90% homology with a peptide sequence capable of being encoded by at least a functional part of the reference nucleotide sequence according to claim 1.

3. Nucleic material of the retroviral genomic type according to either of claims 1 and 2, comprising a nucleic fragment inserted between two sequences corresponding respectively to the LTR region and to the gag gene for the retroviral genomic structure, in particular a nucleic fragment consisting of or comprising the sequence SEQ ID NO: 12.

4. Nucleic material of the subgenomic retroviral type, consisting of a nucleotide sequence identical to SEQ ID NO: 11, with at least one deletion, such as a sequence chosen from SEQ ID NOs: 7 to 9.

5. Nucleic material according to either of claims 1 and 4, comprising at least one functional nucleotide sequence encoding at least one retroviral protein.

6. Nucleic material according to either of claims 1 and 4, comprising at least one regulatory nucleotide sequence.

7. Nucleotide fragment of at least 100 bases, comprising a nucleotide sequence chosen from the group comprising:

a) all the nucleotide sequences, partial and complete, of a nucleic material according to any one of claims 1 to 6

b) all the nucleotide sequences, partial and complete, of a clone chosen from the group including the clones:

- cl.6A2 (SEQ ID NO: 1)
- cl.6A1 (SEQ ID NO: 2)
- cl.7A16 (SEQ ID NO: 3)
- cl.Pi22 (SEQ ID NO: 4)
- cl.24.4 (SEQ ID NO: 5)
- cl.C4C5 (SEQ ID NO: 6)
- cl.PH74 (SEQ ID NO: 7)
- cl.PH7 (SEQ ID NO: 8)
- cl.Pi5T (SEQ ID NO: 9)
- cl.44.4 (SEQ ID NO: 10)
- HERV-W (SEQ ID NO: 11)
- cl.6A5 (SEQ ID NO: 12)
- cl.7A20 (SEQ ID NO: 13)
- cl.7A21 (SEQ ID NO: 14)
- LTR (SEQ ID NO: 15)

c) the sequences which are respectively complementary to the sequences according to a) and b)

d) the sequences which are respectively equivalent to the sequences according to a) to c), in particular the nucleotide sequences exhibiting, for any sequence of 100 contiguous monomers, at least 50%, and preferably at least 70%, for example at least 90% homology with the sequences a) to c).

8. Nucleic probe for the detection of a nucleic material, inserted or otherwise into a nucleic acid, characterized in that it is capable of hybridizing

specifically with a nucleic material, according to any one of claims 1 to 6, or a nucleic fragment according to claim 7.

9. Probe according to claim 8, characterized in that it comprises a marker.

10. Nucleic primer for the amplification by polymerization of an RNA or of a DNA, characterized in that it comprises a nucleotide sequence capable of hybridizing specifically with a nucleic material according to any one of claims 1 to 6, or a nucleic fragment according to claim 7.

11. Nucleic probe or nucleic primer, characterized in that it consists of a nucleotide sequence chosen from the group including SEQ ID NOs: 16 to 28.

12. RNA or DNA, and in particular replication vector, comprising a nucleotide fragment according to claim 7.

13. Peptide encoded by any open reading frame belonging to a nucleotide fragment, according to claim 7, in particular polypeptide, for example oligopeptide forming an antigenic determinant recognized by sera from patients affected by an autoimmune disease, or a pathology which is associated with it, or from patients having a pathological pregnancy or an unsuccessful pregnancy.

14. Peptide according to claim 13, characterized in that it is encoded by a nucleotide fragment comprising an open reading frame encoding one or more retroviral ENV proteins.

15 Use of a nucleic material according to claims 1 to 6, or of a nucleotide fragment according to claim 7, or of a peptide according to claim 13 or 14, as molecular marker for an autoimmune disease or for a pathology which is associated with it, or for a pathological pregnancy or for an unsuccessful pregnancy.

16. Use of a nucleic material according to claims 1 to 6, or of a nucleotide fragment according to claim 7,

as chromosomal marker for susceptibility to an autoimmune disease or for a pathology which is associated with it, or for a risk of a pathological pregnancy or of an unsuccessful pregnancy.

17. Use of a nucleic material according to claims 1 to 6, or of a nucleotide fragment according to claim 7, as proximity marker for a gene for susceptibility to an autoimmune disease or to a pathology which is associated with it, or to a risk of a pathological pregnancy or of an unsuccessful pregnancy.

18. Method for the molecular labeling of an autoimmune disease or of a pathology which is associated with it, of a pathological pregnancy or of an unsuccessful pregnancy, characterized in that any nucleotide fragment according to claim 7, either in RNA form or in DNA form, is identified and/or quantified in any biological body material, in particular body fluid.

19. Method according to claim 18, characterized in that cells expressing the nucleotide fragment according to the claim are detected in said biological body material.

20. Diagnostic or therapeutic composition comprising a nucleic material according to claims 1 to 6, or a nucleotide fragment according to claim 7, or a peptide according to claim 13 or 14.

FIG 1

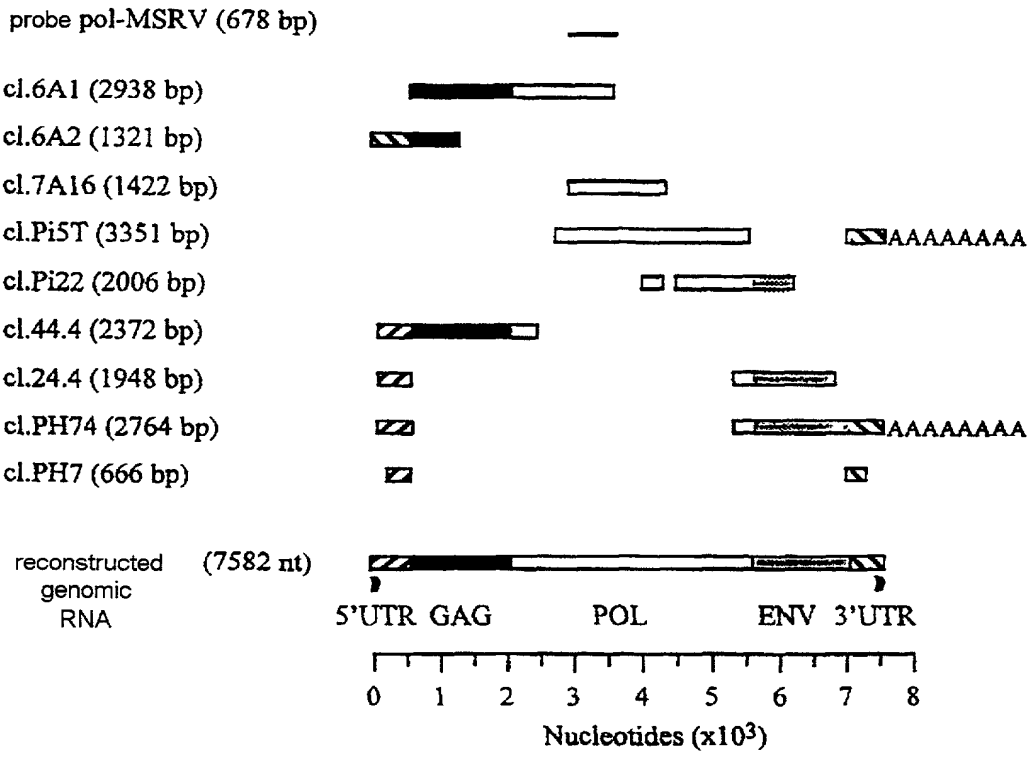


FIG 2

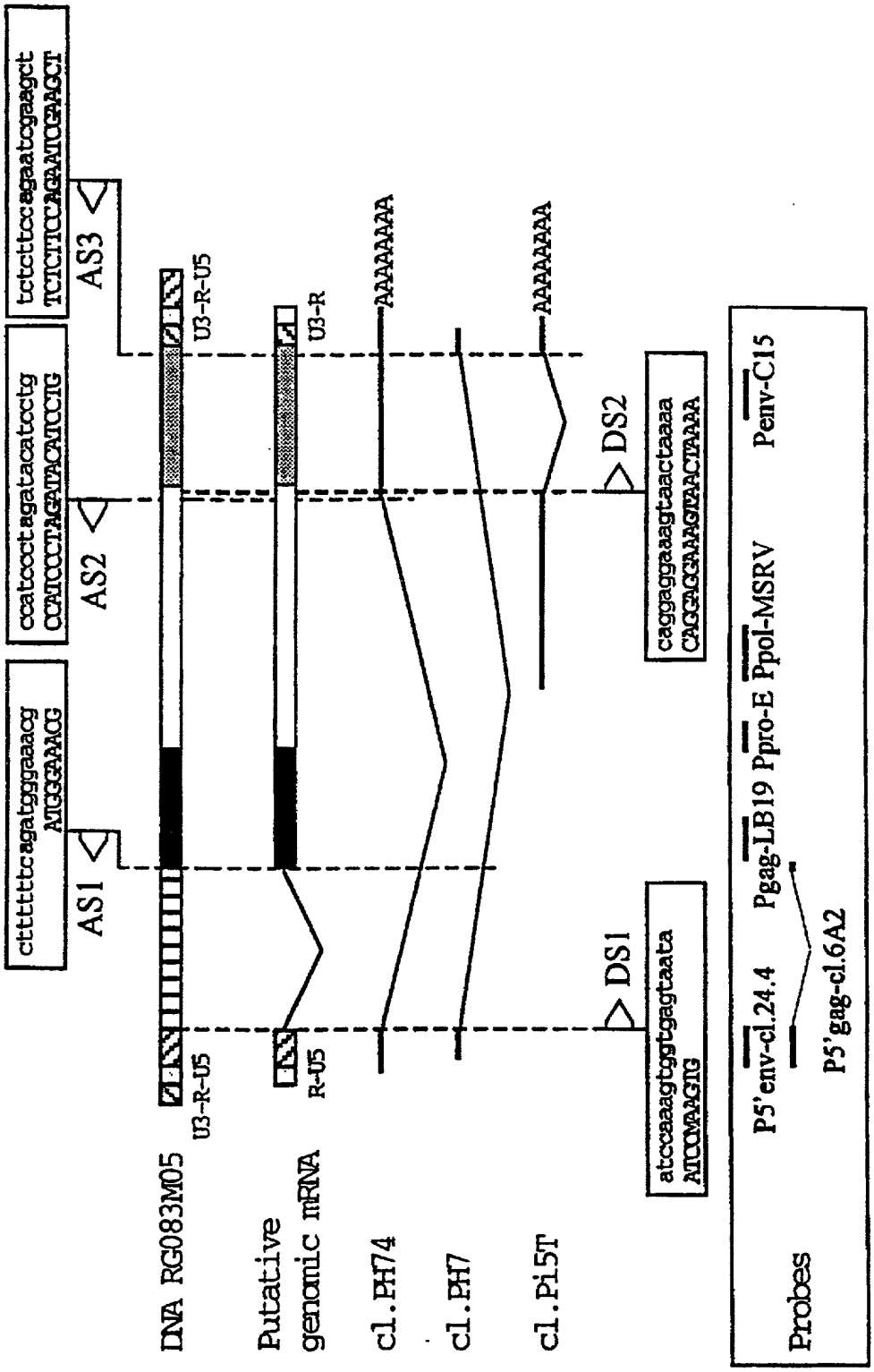
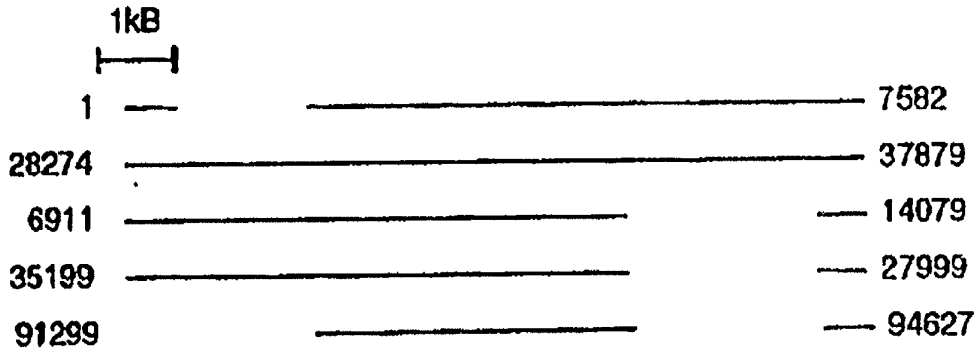

ctttttttcagatgggaaacg
ATGGGAAACG
AS1
ccatcoctagatacatcctg
CCATCCCTAGATACATCCTG
AS2
tctctttccagaatcgaagct
TCTCTTCCAGAATCGAAGCT
AS3
U3-R-U5
U3-R
AAAAAAA
AAAAAAA
DS1
DS2
atccaagtggtgagtaata
ATCCAAGTGGTGAGTAATA
caggaggaaagtaactaaaa
CAGGAGGAAAGTAACTAAAA
DNA RG083M05
U3-R-U5
Putative
genomic mRNA
R-U5
cl.PH74
cl.PH7
cl.Pi5T
Probes
P5'env-cl.24.4
Pgag-LB19
Ppro-E
Ppol-MSRV
Penv-C15
P5'gag-cl.6A2

FIG 3



Names	Similarities	Repetitions	ORFs
Recons RNA		yes	538
RG083M05 [7]	96%	yes	538
BAC378 [14]	88%	yes	no
Q11M15 [21]	89%	yes	413 and 305
U134E6 [x]	88%	no	no

FIG 4A

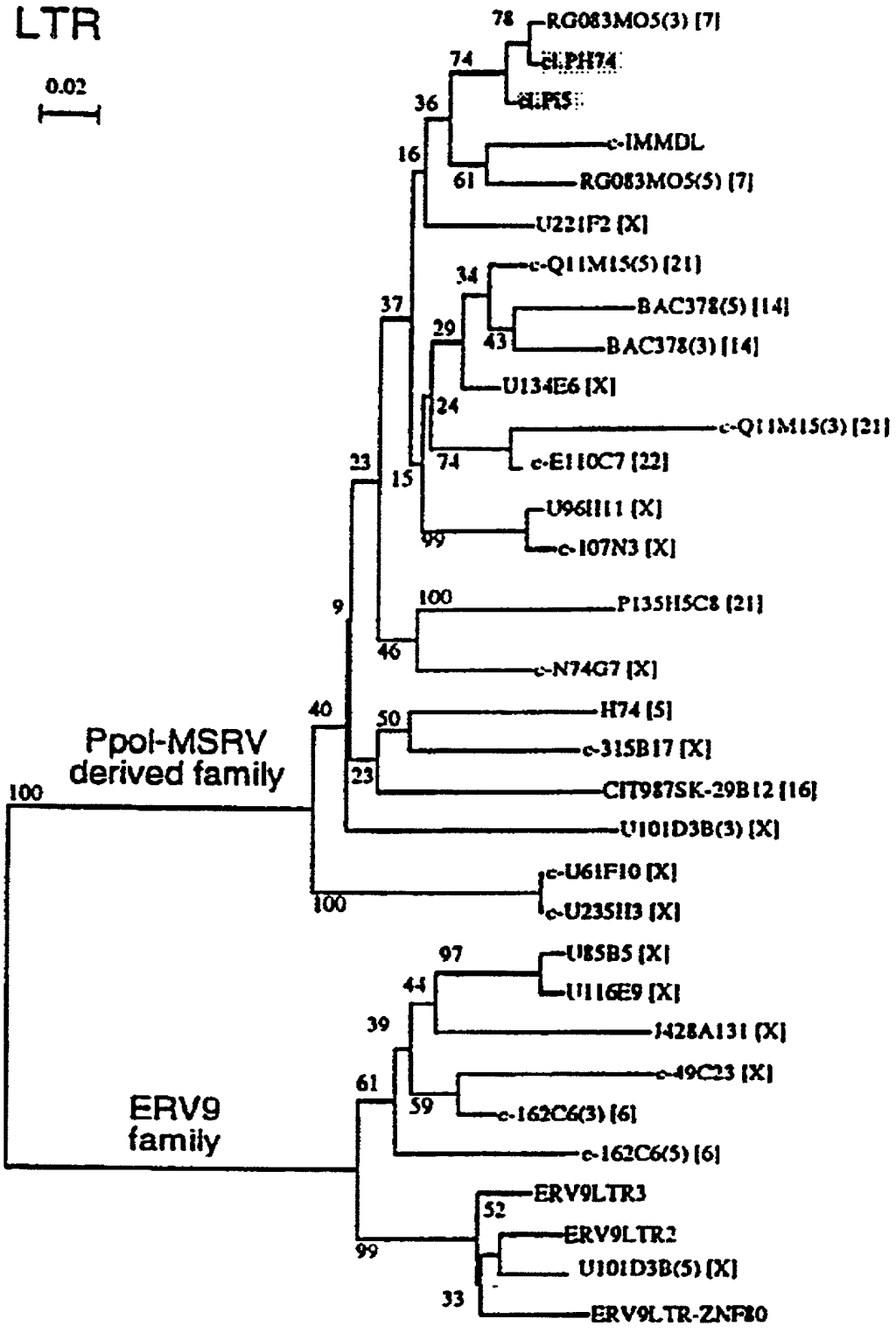

LTR
0.02
RG083MO5(3) [7]
c-IMMDL
RG083MO5(5) [7]
U221F2 [X]
c-Q11M15(5) [21]
BAC378(5) [14]
BAC378(3) [14]
U134E6 [X]
c-Q11M15(3) [21]
c-E110C7 [22]
U96III1 [X]
c-107N3 [X]
P135H5C8 [21]
c-N74G7 [X]
H74 [5]
c-315B17 [X]
CIT987SK-29B12 [16]
U101D3B(3) [X]
c-U61F10 [X]
c-U235H3 [X]
U85B5 [X]
U116E9 [X]
J428A131 [X]
c-49C23 [X]
c-162C6(3) [6]
c-162C6(5) [6]
ERV9LTR3
ERV9LTR2
U101D3B(5) [X]
ERV9LTR-ZNF80
Ppol-MSRV derived family
ERV9 family

FIG 4 B

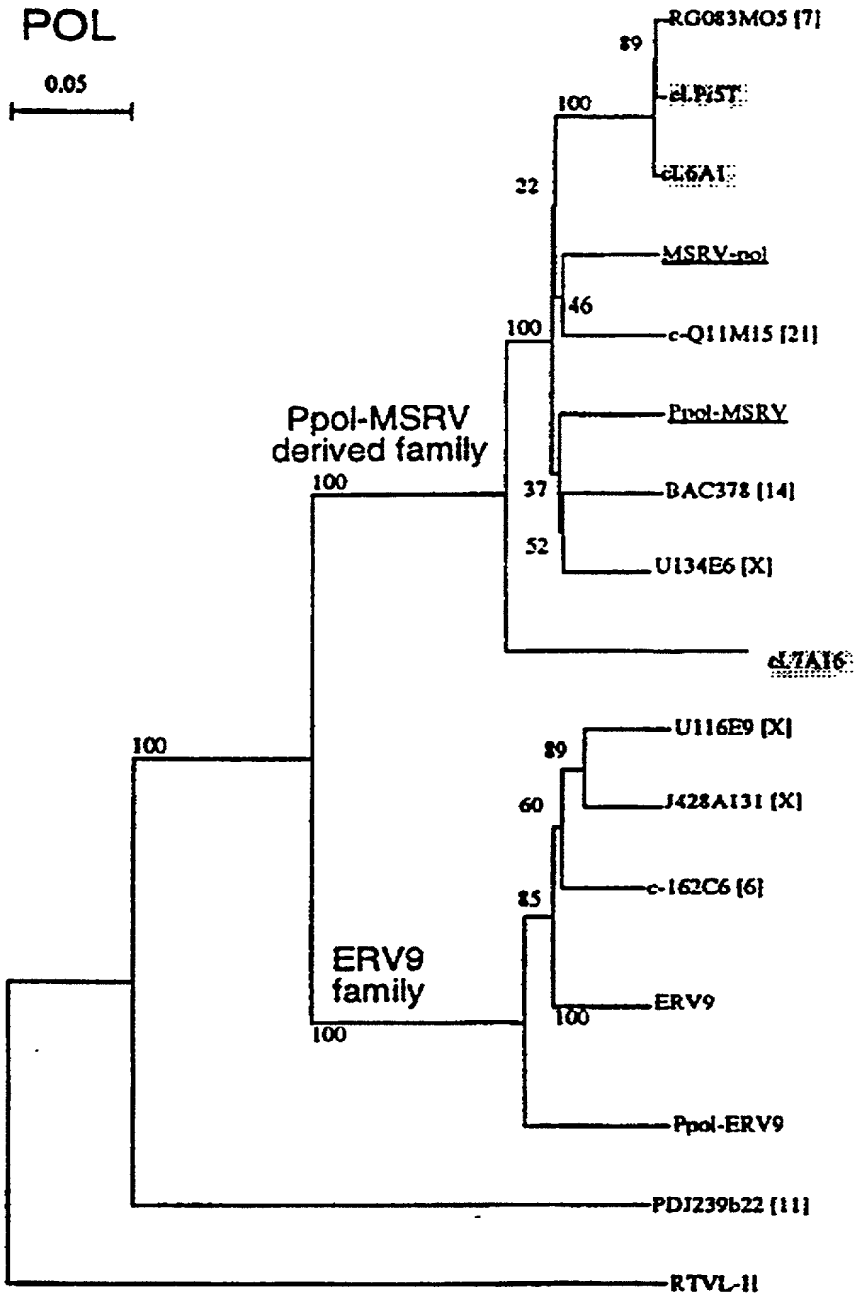

POL
0.05
RG083M05 [7]
89
cLPi5T
100
cL6A1
22
MSRV-pol
46
c-Q11M15 [21]
100
Ppol-MSRV
Ppol-MSRV
derived family
100
37
BAC378 [14]
52
U134E6 [X]
cL7A16
U116E9 [X]
100
89
J428A131 [X]
60
c-162C6 [6]
85
ERV9
family
ERV9
100
100
Ppol-ERV9
PDJ239b22 [11]
RTVL-11

FIG 4C

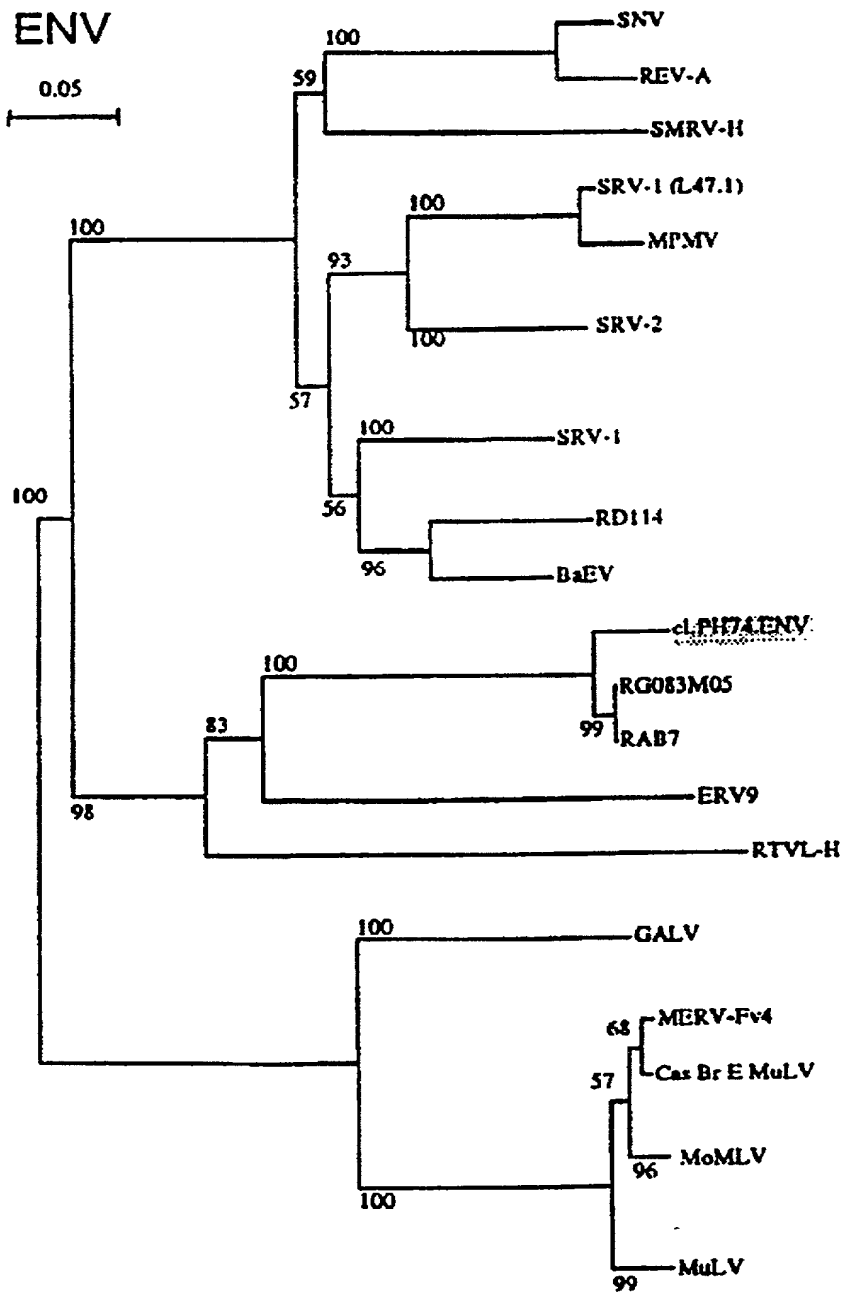

ENV
0.05
SNV
REV-A
SMRV-H
SRV-1 (L47.1)
MPMV
SRV-2
SRV-1
RD114
BaEV
cLPH74ENV
RG083M05
RAB7
ERV9
RTVL-H
GALV
MERV-Fv4
Cas Br E MuLV
MoMLV
MuLV
100
59
100
100
93
100
57
100
56
96
100
98
83
100
99
100
68
57
96
100
99

FIG 5A

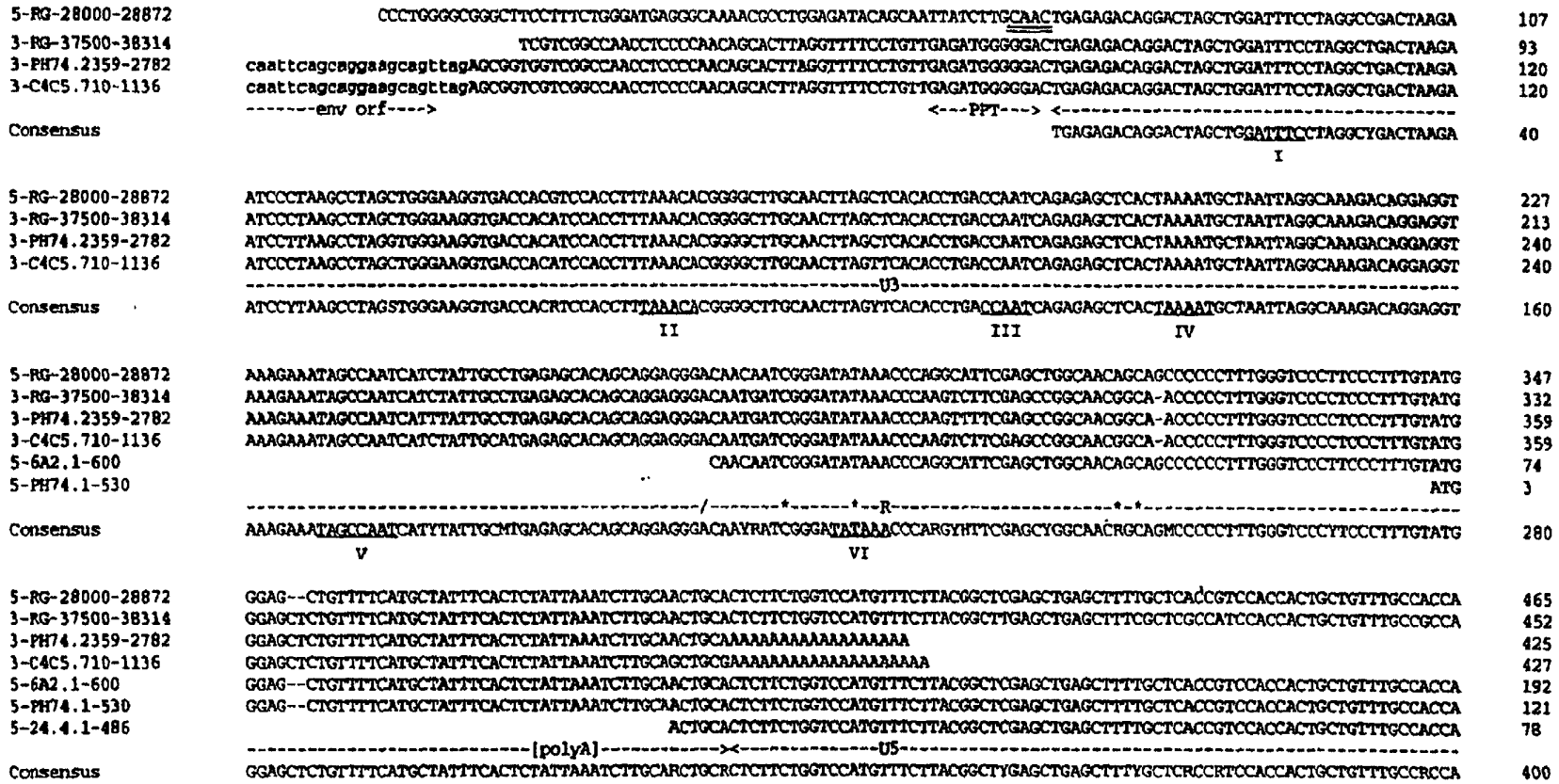


FIG 5B

```
5-RG-28000-28872    CCGCAGACCTGCCGCTGACTCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGCGCCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  585
3-RG-37500-38314    CCGCAGACCCGCCGCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGCACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGC  572
5-6A2.1-600         CCGCAGACCTGCCGCTGACTCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAAGCGCCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  312
5-PH74.1-530        CCGCAGACCTGCCGCTGACTCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAAGCGCCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  241
5-24.4.1-486        CCGCANACCTGCCGCTGACTCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGCGCCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTNCCTGC  198
                    ----------------------------------------------------------U5----------------------------------------------------------
Consensus           CCGCANACCYGCCGCTGACTCCCATCCCTCTGGATCMTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGARGCRCCCATTGCCGCTCCCAATYGGGCTAAAGGCTTGCCATTGTNCCTGC  520

5-RG-28000-28872    ACGGCTAAGTGCCTGGGTTTGTTCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACATGGCCCAAGATT  705
3-RG-37500-38314    ATGGCTAAGTGCCTGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTT  692
5-6A2.1-600         ACGGCTAAGTGCCTGGGTTTGTTCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACATGGCCCAAGATT  432
5-PH74.1-530        ACGGCTAAGTGCCTGGGTTTGTTCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACATGGCCCAAGATT  361
5-24.4.1-486        ACGGCTAAGTGCCTGGGTTTGTTCTAATTGAGCTGAACACTANTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACATGGCCCAAGATT  318
                    ----------------------------------------------------------U5----------------------------------------------------------
Consensus           AYGGCTAAGTGCCTGGGTTYRTYCTAATTGAGCTGAACACTANTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACRGCTTCTAATAGARCTATAACACTYACCRCATGGCCCAAGRTT  640

5-RG-28000-28872    CCATTCCTTGGAATCCGTGAGGCCAA-GAACTCCAGGTCAGAGAATACGAGGCTTGCCACCATCTTGGAAGCGGCCTGCTACCATCTTGGAAGTGGTTCACCACCATCTTGGGAGCTCTG  824
3-RG-37500-38314    CCATTCCTTG-AATCCATAAGGCCAA-GAACCCCAGGTCAGAGAACACGAGGCTTG----------------------------------------CCACCATCTTGGGAGCTCTG  766
5-6A2.1-600         CCATTCCTTGGAATCCGTGAGGCCAA-GAACTCCAGGTCAGAGAATACGAAGCTTGCCACCATCTTGGAAGCGGCCTGCTACCATCTTGGAAGTGGTTCACCACCATCTTGGGAGCTCTG  551
5-PH74.1-530        CCATTCCTTGGAATCCGTGARGCCAACGAACTCCASGTCAGAGAATACGAAGCTTGCCACCATCTTGGAAGCGGCCTGCTACCATCTTGGAAGTGGTTCACCACCATCTTGGGAGCTCTG  481
5-24.4.1-486        CCATTCCTTGGAATCCGTGAGGGCAA-GAACTCCAGGTCAGAGAATACGAGGCTTGCCACCATCTTGGAAGCGGCCTGCTACCATCTTGGAAGTGGTTCACCACCATCTTGGGAGCTCTG  437
                    ----------------------------------------------------------U5----------------------------------------------------------
Consensus           CCATTCCTTGGAATCCRTRARGSCAACGAACYCCASGTCAGAGAAYACGARGCTTGCCACCATCTTGGAAGCGGCCTGCTACCATCTTGGAAGTGGTTCACCACCATCTTGGGAGCTCTG  760

5-RG-28000-28872    TGAGCAAGGACCCCCCGGTAACATTTTGGCAACCACGAACGGACATCCA  873
3-RG-37500-38314    TGAGCAAGGACCCCCAAGTAACACAACCATGAGGGTGCAAATGCATGGG  815
5-6A2.1-600         TGAGCAAGGACCCCCCGGTAACATTTTGGCAACCACGAACGGACATCCA  600
5-PH74.1-530        TGAGCAAGGACCCCCCGGTGACATTTTGGCGACCACCAACGGACATCCC  530
5-24.4.1-486        TGAGCAAGGACCCCCCGGTAACATTTTGGCAACCACGAACGGACATCCA  486
                    ---------------------->  <-------PBS------>
Consensus           TGAGCAAGGACCCCCMRGTRACA  783
```

ORF1: ENV (538 AA) FIG6

```
<---            L        --->---<---      SU
MGLPYHIFLCSVLSPCFTLTAPPPCRCMTSSSPHPEFLWRMQRPGNIDAPSYRSLSKGTP     60
 A          FT V   S                YQ                     C

TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV    120

                                                      •
QDQAREKHVKEVISQLTGVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA    180
                R
•                         •     •                          •
QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNVEITHTSNLTCVKF    240
                                                   L
•                                             •
SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMAIY    300
                                                           T
              --->---<---   TM
TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER    360

                                                      •
VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVEE    420
                   R     S                                 K

IPDRIQRIAEELRNTGPWGLLSRWMPWILPFLGPLAAIILLLLFGPCIFDLLVNFVSSRI    480
 R    R                Q                      N
                                                         --->
EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSAGSS      538
```

ORF2 (52AA)

MEPKMQSKTKIYRRPLDRPVSPRSDVNDIKGTPPEEISAAQPLLRPNSAGSS-

Alignment ORF2 and Rex PLLV-L

```
ORF2                   KIY-RRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRP
                       ++Y    LD P SP  ++        P   S  QPLLRP
Rex PTLV-L (B53482)    RLYNTLSLDSPPSPPKELPA-----PSRFSPPQPLLRP
```

ORF3 (48AA)

MLMTSKAPLLRKSQLHNLYYAPIQQEAVRAVVGQPPQQHLGFPVEMGD

Alignment ORF3 and Tat SIV-AGM

```
ORF3                   MTSKAPLLRKSQLHNLYYAPIQQEAVRAVVGQPPQ
                       +T  AP  R+ ++ +L  AP+Q +++    G+  Q
Tat SIV-AGM(p05913)    VTYHAPRTRRKKIRSLNLAPLQHQSISTKWGRDGQ
```

DECLARATION AND POWER OF ATTORNEY UNDER 35 USC §371(c)(4) FOR PCT APPLICATION FOR UNITED STATES PATENT

As a below named inventor, I hereby declare that:

my residence, post office address and citizenship are as stated below under my name;

I verily believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought, namely the invention entitled: Endogenetic retroviral sequences, associated with autoimmune diseases or with pregnancy disorders

described and claimed in international application number _____ filed _____.

I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations §1.56. Under Title 35, U.S. Code §119, the priority benefits of the following foreign application(s) filed within one year prior to my international application are hereby claimed:

French patent application No 97 08815 filed July 7, 1997

The following application(s) for patent or inventor's certificate on this invention were filed in countries foreign to the United States of America either (a) more than one year prior to my international application, or (b) before the filing date of the above-named foreign priority application(s):

I hereby appoint the following as my attorneys of record with full power of substitution and revocation to prosecute this application and to transact all business in the Patent Office:

James A. Oliff, Reg. No. 27,075; William P. Berridge, Reg. No. 30,024; Kirk M. Hudson, Reg. No. 27,562; Thomas J. Pardini, Reg. No. 30,411; and Edward P. Walker, Reg. No. 31,450.

5

ALL CORRESPONDENCE IN CONNECTION WITH THIS APPLICATION SHOULD BE SENT TO OLIFF & BERRIDGE, P.O. BOX 19928, ALEXANDRIA, VIRGINIA 22320, TELEPHONE (703) 836-6400.

I hereby declare that I have reviewed and understand the contents of this Declaration, and that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Typewritten Full Name of Sole or First Inventor: 100 Frédéric (Given Name) _____ (Middle Initial) BESEME (Family Name)

Inventor's Signature: Frédéric Beseme

Date of Signature: November 8, 1999.

Residence: Villefontaine (City) _____ (State or Province) FRANCE FRX (Country)

Citizenship: French

Post Office Address (Insert complete mailing address, including country): 39 rue de la Noyera, 38090 Villefontaine, FRANCE

Note to Inventor: Please sign name on line 2 exactly as it appears in line 1 and insert the actual date of signing on line 3.

IF THERE IS MORE THAN ONE INVENTOR USE PAGE 2 AND PLACE AN "X" HERE X

(Discard this page in a sole inventor application)

Typewritten Full Name of Joint Inventor 2-00
Jean-Luc (Given Name) — (Middle Initial) — BLOND (Family Name)
Inventor's Signature Jean-Luc Blond
Date of Signature November 8, 1999
Residence LYON (City) — (State or Province) — FRANCE FRX (Country)
Citizenship French
Post Office Address (Insert complete mailing address, including country) 75 bis rue des Acqueducs, 69005 Lyon, FRANCE

Typewritten Full Name of Joint Inventor 3-00
Olivier (Given Name) — (Middle Initial) — BOUTON (Family Name)
Inventor's Signature Olivier BOUTON
Date of Signature November 8, 1999
Residence Francheville (City) — (State or Province) — FRANCE FRX (Country)
Citizenship French
Post Office Address (Insert complete mailing address, including country) 48 Avenue du Châter, 69340 Francheville, FRANCE

Typewritten Full Name of Joint Inventor 4-00
Bernard (Given Name) — F. (Middle Initial) — MANDRAND (Family Name)
Inventor's Signature Bernard F. Mandrand
Date of Signature November 8, 1999
Residence Villeurbanne (City) — (State or Province) — FRANCE FRX (Country)
Citizenship French
Post Office Address (Insert complete mailing address, including country) 21 rue de la Doua, 69100 Villeurbanne, FRANCE

Typewritten Full Name of Joint Inventor 5-00
François (Given Name) — (Middle Initial) — MALLET (Family Name)
Inventor's Signature François Mallet
Date of Signature November 8, 1999
Residence Villeurbanne (City) — (State or Province) — FRANCE FRX (Country)
Citizenship French
Post Office Address (Insert complete mailing address, including country) 84 rue Anatole France, 69100 Villeurbanne, FRANCE

Typewritten Full Name of Joint Inventor 6-00
Hervé (Given Name) — (Middle Initial) — PERRON (Family Name)
Inventor's Signature Hervé PERRON
Date of Signature November 8, 1999
Residence Lyon (City) — (State or Province) — FRANCE FRX (Country)
Citizenship French
Post Office Address (Insert complete mailing address, including country) 15 rue de Boyer, 69005 Lyon, FRANCE

Note to Inventor: Please sign name on line 2 exactly as it appears in line 1 and insert the actual date of signing on line 3.

This form may be executed only when attached to the first page of the Declaration and Power of Attorney of the application to which it pertains.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: BIO MERIEUX
(B) STREET: CHEMIN DE L'ORME
(C) CITY: MARCY L'ETOILE
(E) COUNTRY: FRANCE
(F) POSTAL CODE: 69280

(ii) TITLE OF INVENTION: NUCLEIC MATERIAL OF THE ENDOGENOUS RETROVIRAL GENOMIC TYPE, ASSOCIATED WITH AN AUTOIMMUNE DISEASE AND/OR WITH PREGNANCY DISORDERS; USE AS MARKER

(iii) NUMBER OF SEQUENCES: 35

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1321 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CAACAATCGG GATATAAACC CAGGCATTCG AGCTGGCAAC AGCAGCCCCC CTTTGGGTCC 60

CTTCCCTTTG TATGGGAGCT GTTTTCATGC TATTTCACTC TATTAAATCT TGCAACTGCA 120

CTCTTCTGGT CCATGTTTCT TACGGCTCGA GCTGAGCTTT TGCTCACCGT CCACCACTGC 180

TGTTTGCCAC CACCGCAGAC CTGCCGCTGA CTCCCATCCC TCTGGATCCT GCAGGGTGTC 240

CGCTGTGCTC CTGATCCAGC GAAGCGCCCA TTGCCGCTCC CAATTGGGCT AAAGGCTTGC 300

CATTGTTCCT GCACGGCTAA GTGCCTGGGT TTGTTCTAAT TGAGCTGAAC ACTAGTCACT 360

GGGTTCCATG GTTCTCTTCT GTGACCCACG GCTTCTAATA GAACTATAAC ACTTACCACA 420

TGGCCCAAGA TTCCATTCCT TGGAATCCGT GAGGCCAAGA ACTCCAGGTC AGAGAATACG 480

AAGCTTGCCA CCATCTTGGA AGCGGCCTGC TACCATCTTG GAAGTGGTTC ACCACCATCT 540

TGGGAGCTCT GTGAGCAAGG ACCCCCCGGT AACATTTTGG CAACCACGAA CGGACATCCA 600

AAGTGATGGG AAACGTTCCC CGCAAGACAA AAACGCCCCT AAGACGTATT CTGGAAAATT 660

GGGAACAATT TGACCCTCAG ACACTAAGAA AGAAACGACT TATATTCTTC TGCAGTGCCG 720

CCTGGCACTC CTGAGGGAAG TATAAATTAT AACACCATCT TACAGCTAGA CCTCTTTTGT 780

AGAAAAGGCA AATGGAGTGA AGTGCCATAA GTACAAACTT TCTTTTCATT AAGAGACAAC 840

TCACAATTAT GTAAAAAGTG TGATTTATGC CCTACAGGAA GCCTTCAGAG TCTACCTCCC 900

TATCCCAGCA TCCCCGACTC CTTCCCCACT TAATAAGGAC CCCCCTTCAA CCCAAATGGT 960

CCAAAAGGAG ATAGACAAAA GGGTAAACAG TGAACCAAAG AGTGCCAATA TTCCCCAATT 1020

ATGACCCCTC CAAGCAGTGG GAGGAAGAGA ATTCGGCCCA GCCAGAGTGC ATGTGCCTTT 1080

TTCTCTCCCA GACTTAAAGC AAATAAAAAC AGACTTAGGT AAATTCTCAG ATAACCCTGA 1140

TGGCTATATT GGTGTTTTAC AAGGGTTAGG ACAATTCTTT GATCTGACAT GGAGAGATAT 1200

ATATGTCACT GCTAAATCAG ACACTAACCC CAAATGAGAG AAGTGCCACC ATAACTGCAG 1260

CCTGAGAGTT TGGCGATCTC TGGTATCTCA GTCAGGTCAA TGATAGGATG ACAACAGAGG 1320

A 1321

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2938 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CAACGACGGA CATCCAAAGT GATGGGAAAC GTTCCCCGCA AGACAAAAAC GCCCCTAAGA 60

CGTATTCTGG AGAATTGGGA CCAATTTGAC CCTCAGACAC TAAGAAAGAA ACGACTTATA 120

TTCTTCTGCA GTGCCGCCTG GCACTCCTGA GGGAAGTATA AATTATAACA CCATCTTACA 180

GCTAGACTTC TTTTGTAGAA AAGGCAAATG GAGTGAAGTG CCATAAGTAC AAACTTTCTT 240

TTCATTAAGA GACAACTCAC AATTATGTAA AAAGTGTGAT TTATGCCCTA CAGGAAGCCT 300

TCAGAGTCTA CCTCCCTATC CCAGCATCCC CGACTCCTTC CCCAACTAAT AAGGACCCCC 360

CTTCAACCCA AATGGTCCAA AAGGAGATAG ACAAAAGGGT AAACAGTGAA CCAAAGAGTG 420

CCAATATTCC CCAATTATGA CCCCTCCCAA GCAGTGGGAG GAAGAGATTC GGCCCAGCCA 480

GAGTGCATGT GCTTTTTCTT CTCCCAGACT TAAAGCAAAT AAAAACAGAC TTAGGTAAAT 540

TCTCAGATAA TCCTGATGGC TATATTGATG TTTTACAAGG GTTAGGACAA TTCTTTGATC 600

TGACATGGAG AGATATAATG TCACTGCTAA ATCAGACACT AACCCCAAAT GAGAGAAGTG 660

CCACCATAAC TGCAGCCTGA GAGTTTGGCG ATCTCTGGTA TCTCAGTCAG GTCAATGATA 720

GGATGACAAC AGAGGAAAGA GATGATCCCC ACAGCCAGCA AGCAGTTCCC AGTCTASACC 780

CTCATTGGGG ACACAGAAAT CAGTAACATG GGAGATTGGT GCTGCAGACA TTTGCTAACT 840

TGTGTGCTAC AAGGACTAAG GAAAACTACG AAGAAAATCT ACGAATTACT CAATGATGTC 900

CACCATAACA CAGGGGAAGG GAAGAAAATC CTACTGCCTT TCTGGAGAGA CTAAGGGAGG 960

CATTGAGGAA GCGTGCCTCT CTGTCACCTG ACTCTTCTGA AGGCCAACTA ATCTTAAAGC 1020

GTAAGTTTAT CACTCAGTCA GCTGCAGACA TTAGAAAAAA CTTCAAAAGT CTGCCGTAGG 1080

CCCGGAGCAA AACTTAGAAA CCCTATTGAA CTTGGCAACY TCGGTTTTTT ATAATAGAGA 1140

TCAGGAGGAG CAGGCGGAAC AGGACAAACG GGATTAAAAA AAAGGCCACC GCTTTAGTCA 1200

TGACCCTCAG GCAAGTGGAC TTTGGAGGCT CTGGAAAAGG GAAAAGCTGG GCAAATTGAA 1260

TGCCTAATAG GGCTTGCTTC CAGTGCGGTC TACAAGGACA CTTTAAAAAA GATTGTCCAA 1320

GTAGAAGTAA GCCGCCCCTT CGTCCATGCC CCTTATTTCA AGGGAATCAC TGGAAGGCCC 1380

ACTGCCCCAG GGGACAAAGG TCTTTTGAGT CAGAAGCCAC TAACCAGATG ATCCAGCAGC 1440

AGGACTGAGG GTGCCTGGGG CAAGCGCCAT CCCATGCCAT CACCCTCACA GAGCCCTGGG 1500

TATGCTTGAC CATTGAGGGC CAGGAAGGTT GTCTCCTGGA CACTGGTGCG GTCTTCTTAG 1560

TCTTACTCTT CTGTCCCGGA CAACTGTCCT CCAGATCTGT CACTATCTGA GGGGGTCCTA 1620

AGACGGGCAG TCACTAGATA CTTCTCCCAG CCACTAAGTT ATGACTGGGG AGCTTTATTC 1680

TTTTCACATG CTTTTCTAAT TATGCTTGAA AGCCCCACTA CCTTGTTAGG GAGAGACATT 1740

CTAGCAAAAG CAGGGGCCAT TATACACCTG AACATAGGAG AAGGAACACC CGTTTGTTGT 1800

CCCCTGCTTG AGGAAGGAAT TAATCCTGAA GTCTGGGCAA CAGAAGGACA ATATGGACGA 1860

GCAAAGAATG CCCGTCCTGT TCAAGTTAAA CTAAAGGATT CCACTTCCTT TCCCTACCAA 1920

AGGCAGTACC CCCTCAGACC CAAGGCCCAA CAAGGATTCC AAAAGATTGT TAAGGACTTA 1980

AAAGCCCAAG GCTTAGTAAA ACCATGCATA ACTCCCTGCA GTAATTCCGT AGTGGATTGA 2040

GGAGGCACAG AAACCCAGTG GACAGTGGAG GGTTAGTGCA AGATCTCAGG ATTATCAATG 2100

GAGGCCGTTG TCCTTTTATA CCCAGCTGTA CCTAGCCCTT ATACTGTGCT TTCCCAAATA 2160

CCAGAGGAAG CAGAGTGGTT TACACTCCTG GACCTTAAGG ATGCCTTCTT CTGCATCCCT 2220

GTACATCCTG ACTCTCAATT CTTGTTTGCC TTTGAAGATA CTTCAAACCC AACATCTCAA 2280

CTCACCTGGA CTGTTTTACC CCAAGGGTTC AGGGATAGCC CCCATCTATT TGGCCAGGCA 2340

TTAGCCCAAG ACTTGAGCCA ATCCTCATAC CTGGACACTT GTCCTTCGGT AGGTGGATGA 2400

TTTACTTTTG GCCGCCCATT CAGAAACCTT GTGCCATCAA GCCACCCAAG CGCTCTTCAA 2460

TTTCCTCGCT ACCTGTGGCT ACATGGTTTC CAAACCAAAG GCTCAACTCT GCTCACAGCA 2520

GGTTACTTAG GGCTAAAATT ATCCAAAGGC ACCAGGGCCC TCAGTGAGGA ACACATCCAG 2580

CCTATACTGG CTTATCCTCA TCCCAAAACC CTAAAGCAAC TAAGGGGATT CCTTGGCGTA 2640

ATAGGTTTCT GCCGAAAATG GATTCCCAGG TTTGGCGAAA TAGCCAGGTC ATTAAATACA 2700

CTAATTAAGG AAACTCAGAA AGCCAATACC CATTTAGTAA GATGGACAAC TGAAGTAGAA 2760

GTGGCTTTCC AGGCCCTAAC CCAAGCCCCA GTGTTAAGTT TGCCAACAGG GCAAGACTTT 2820

TCTTCATATG TCACAGAAAA AACAGGAATA GCTCTAGGAG TCCTTACACA GATCCGAGGG 2880

ATGAGCTTGC AACCTGTGGC GTACCTGACT AAGGAAATTG ATGTAGTGGC AAAGGGTT 2938

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1422 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TCAGGGATAG CCCCCATCTA TTTGGCCAGG CATTAGCCCA AGACTTGAGT CAGTTATCAT 60

ACCTGGACAC TCTTGTCCTT CAGTATGTGG ATGATTTACT TTTAGCTGCC TGTTCAGAAA 120

CCTTGTGCCA TCAAGCCACC CAAGCACTCT TAAATTTCCT CGCCACCTGT GGCTACAAGG 180

TTTCCAAAGA GAAGCTCAGC TCTGCTCACA GCAGGTTAAA TACTTAGGAC TAAGATTATC 240

CAAAGGCACC AAGGCCCTCA GTGAGGAATG TATCCAGCCT ATACTGGCTT ATCCTCATCT 300

CAAAACCCTA AAGCAACTAA GAGAGTTCCT TGGCATAACA GGCTTCTGCC GAATATGGAT 360

TCCCCAGGTA TGGCAAAATA GCCAGGCCAT TATATACAGT AATTAAGGAA ACTCAGAAAG 420

CCAATACCCA TTTAATAAGA TGGATACCTG AAGCCAAAGT GGCTTTCCAG GCCCCTAAAG 480

AAGGCCTTAA ACCCAAGTCC CAGTGTTAAG CTTGCCAACG GGGCAAGACT TTTCTTTATA 540

CATCACAGAA AAAAACAGAA ACAGCTCTGG GAGTCCTTAC ACAGGTCCAA GGGACGAGCT 600

TGCAACCCAT GGCATACCTG AGTAAGGAAA CTGATGTAGT GGCAAAGGGT TGGCTTCATT 660

GTTTATGGGT AGTGGTGGCA GTAGCAGTTG TAGTATCTGA AGCAGTTAAA ATAATACAGG 720

GGAGAGATCT TACTGTGTGG ACATCTCATG AGGTGAACAG CATACTCACT GCTAAAGGAG 780

ACTTGTGGCT GTCAGACAAC CGTTTACTTA AATATCAGGC TCTATTACTT GAAAGGCCAG 840

TGCTGCAACT GTGCACTTGT GCAACTCTTA ACCCAGTCNC ATTTCTTCCA GACAATGAAG 900

ATAGAATATA ACTGTCAACA AATAATTTCT CAAACCTATG CCACTCGAGG GGACCTTCTA 960

GAAGTTCCCT TGACTGATCC TGACCTTCAA CTTGTATACT GATGGAAGTT CCTTTGTAGA 1020

AAAAGGACTT CAAAAGCGGG GTATGCAGTG GTCAGTGATA ATGGAATATT TGAAAGTATC 1080

CCCTCACTCC AGGAACTAGT GCTTAGCTGG CAGAACTAAT AGCCTTCATT GGGGCACTAG 1140

AATTAGGAGA AGGAAAAAGG GTAAATATAT ATACAGACTC TGAGTATGCT CACCTAGTCN 1200

TCCATGCCCA TGAGGCAATA TGCAGAGAAA GGGAATTCCT AACTTCCGAG GGAACACCTA 1260

TCACACATCA GGAAGCCATT AGGAGATTAT TACTGGCAGT ACAGAAACCT AAAGAGGTGG 1320

AAGTCTTACA CTGCTGGGGT CATCAGAAAG GAAAGAAAAG GGAAATAGAA GGGAATTGCC 1380

AAGCAGATAT TGAAGCAAAA AGAGCTGCAA GGCAGGACCC TC 1422

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2006 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATGCAGTGGT CAGTGATAAT GGAATACTTG AAAGTAATCC CCTCACTCCA GGAACTAGTG 60

CTCAGCTAGC AGAACTAATA GCCCTCACTT GGGCACTAGA ATTAGGAGAA GAAAAAAGGG 120

CAAATATATA TACAGACTCT AAATATGCTT ACCTAGTCCT CCATGCCCAT GCAGCAATAT 180

GGAAAGAAAG GGAATTCCTA ACTTCTGAGA GAACACCTAT CAAACATCAG GAAGCCATTA 240

GGAAATTATT ATTGGCTGTA CAGAAACCTA AAGAGGTGGC AGTCTTACAC TGCCGGGGTC 300

ATCANAAAGG AAAGGAAAGG GAAAATACTT TTGCCTGCAA CTATCCAATG GAAATTACTT 360

AAAACCCTTC ATCAAACCTT TCACTTAGGC ATCGATAGCA CCCATCAAAT GGCCAAATCA 420

TTATTTACTG GACCAGGCCT TTTCAAAACT ATCAAGCAAA TATTCAGGGC CTGTGAATTG 480

TGCCAAAAAA ATAATCCCCT GCCTCATCGC CAAGCTCCTT CAGGAAAACA AAAAACAGGC 540

CATTACCCTG AAAAAAACTG GCAACTGATT TTACCCACAA GCCCAAACCT CAGGGATTTC 600

AGTATCTACT AGTCTGGGTA AATACTTTCA CGGGTTGGGC AAAGGCCTTC CCCTGTAGGA 660

CAGAAAAGGC CCAAGAGGTA ATAAAGGCAC TAGTTCATGA AATAATTCCC AGATTCGGAC 720

TTCCCCGAGG CTTACAGAGT GACAATAGCC CTGCTTTCCA GGCCACAGTA ACCCAGGGAG 780

TATCCCAGGC GTTAGGTATA CGATATCACT TACACTGCGC CTGAAGGCCA CAGTCCTCAG 840

GGAAGGTCGA GAAAATGAAT GAAATACTCA AAGGACATCT AAAAAAGCAA ACCCAGGAAA 900

CCCACCTCAC ATGGCCTGCT CTGTTGCCTA TAGCCTTAAA AAGAATCTGC AACTTTCCCC 960

AAAAAGCAGG ACTTAGCCCA TACGAAATGC TGTATGGAAG GCCCTTCATA ACCAATGACC 1020

TTGTGCTTGA CCCAAGACAG CCAACTTAGT TGCAGACATC ACCTCCTTAG CCAAATATCA 1080

ACAAGTTCTT AAAACATTAC AAGGAACCTA TCCCTGAGAA GAGGGAAAAG AACTATTCCA 1140

CCCTTGTGAC ATGGTATTAG TCAAGTCCCT TCTCTCTAAT TCCCCATCCC TAGATACATC 1200

CTGGGAAGGA CCCTACCCAG TCATTTTATT TACCCCAACT GCGGTTAAAG TGGCTGGAGT 1260

GGTCTTGGAT ACATCACACT TGAGTCAAAT CCTGGATACT GCCAAAGGAA CCTGAAAATC 1320

CAGGAGACAA CGCTAGCTAT TCCTGTGAAC CTCTAGAGGA TTTGCGCCTG CTCTTCAAAC 1380

AACAACCAGG AGGAAAGTAA CTAAAATCAT AAATCCCCCA TGGCCCTCCC TTATCATATT 1440

TTTCTCTTTA CTGTTCTTTT ACCCTCTTTC ACTCTCACTG CACCCCCTCC ATGCCGCTGT 1500

ATGACCAGTA GCTCCCCTTA CCAAGAGTTT CTATGGAGAA TGCAGCGTCC CGGAAATATT 1560

GATGCCCCAT CGTATAGGAG TCTTTCTAAG GGAACCCCCA CCTTCACTGC CCACACCCAT 1620

ATGCCCCGCA ACTGCTATCA CTCTGCCACT CTTTGCATGC ATGCAAATAC TCATTATTGG 1680

ACAGGAAAAA TGATTAATCC TAGTTGTCCT GGAGGACTTG GAGTCACTGT CTGTTGGACT 1740

TACTTCACCC AAACTGGTAT GTCTGATGGG GGTGGAGTTC AAGATCAGGC AAGAGAAAAA 1800

CATGTAAAAG AAGTAATCTC CCAACTCACC CGGGTACATG GCACCTCTAG CCCTACAAAG 1860

GACTAGATCT CTCAAAACTA CATGAAACCC TCCGTACCCA TACTCGCCTG GTAAGCCTAT 1920

TTAATACCAC CCTCACTGGG CTCCATGAGG TCTCGGCCCA AAACCCTACT AACTGTTGGA 1980

TATGCCTCCC CCTGAACTTC AAGCCA 2006

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1948 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ACTGCACTCT TCTGGTCCAT GTTTCTTACG GCTCGAGCTG AGCTTTTGCT CACCGTCCAC 60

CACTGCTGTT TGCCACCACC GCANACCTGC CGCTGACTCC CATCCCTCTG GATCCTGCAG 120

GGTGTCCGCT GTGCTCCTGA TCCAGCGAGG CGCCCATTGC CGCTCCCAAT TGGGCTAAAG 180

GCTTGCCATT GTNCCTGCAC GGCTAAGTGC CTGGGTTTGT TCTAATTGAG CTGAACACTA 240

NTCACTGGGT TCCATGGTTC TCTTCTGTGA CCCACGGCTT CTAATAGAAC TATAACACTT 300

ACCACATGGC CCAAGATTCC ATTCCTTGGA ATCCGTGAGG GCAAGAACTC CAGGTCAGAG 360

AATACGAGGC TTGCCACCAT CTTGGAAGCG GCCTGCTACC ATCTTGGAAG TGGTTCACCA 420

CCATCTTGGG AGCTCTGTGA GCAAGGACCC CCCGGTAACA TTTTGGCAAC CACGAACGGA 480

CATCCAAAGT GATACATCCT GGGAAGGACC CTACCCAGTC ATTTTATCTA CCCCAACTGC 540

GGTTAAAGTG GCTGGAGTGG AGTCTTGGAT ACATCACACT TGAGTCAAAT CCTGGATACT 600

GCCAAAGGAA CCTGAAAATC CAGGAGACAA CGCTAGCTAT TCCTGTGAAC CTCTAGAGGA 660

TTTGCGCCTG CTCTTCAAAC AACAACCAGG AGGAAAGTAA CTAAAATCAT AAATCCCCAT 720

GGCCCTCCCT TATCATATTT TTCTCTTTAC TGTTGTTTCA CCCTCTTTCA CTCTCACTGC 780

ACCCCCTCCA TGCCGCTGTA TGACCAGTAG CTCCCCTTAC CAAGAGTTTC TATGGAGAAT 840

GCAGCGTCCC GGAAATATTG ATGCCCCATC GTATAGGAGT CTTTGTAAGG GAACCCCCAC 900

CTTCACTGCC CACACCCATA TGCCCCGCAA CTGCTATCAC TCTGCCACTC TTTGCATGCA 960

TGCAAATACT CATTATTGGA CAGGAAAAAT GATTAATCCT AGTTGTCCTG GAGGACTTGG 1020

AGTCACTGTC TGTTGGACTT ACTTCACCCA AACTGGTATG TCTGATGGGG GTGGAGTTCA 1080

AGATCAGGCA AGAGAAAAAC ATGTAAAAGA AGTAATCTCC CAACTCACCC GGGTACATGG 1140

CACCTCTAGC CCCTACAAAG GACTAGATCT CTCAAAACTA CATGAAACCC TCCGTACCCA 1200

TACTCGCCTG GTAAGCCTAT TTAATACCAC CCTCACTGGG CTCCATGAGG TCTCGGCCCA 1260

AAACCCTACT AACTGTTGGA TATGCCTCCC CCTGAACTTC AGGCCATATG TTTCAATCCC 1320

TGTACCTGAA CAATGGAACA ACTTCAGCAC AGAAATAAAC ACCACTTCCG TTTTAGTAGG 1380

ACCTCTTGTT TCCAATCTGG AAATAACCCA TACCTCAAAC CTCACCTGTG TAAAATTTAG 1440

CAATACTACA TACACAACCA ACTCCCAATG CATCAGGTGG GTAACTCCTC CCACACAAAT 1500

AGTCTGCCTA CCCTCAGGAA TATTTTTTGT CTGTGGTACC TCAGCCTATC GTTGTTTGAA 1560

TGGCTCTTCA GAATCTATGT GCTTCCTCTC ATTCTTAGTG CCCCCTATGG CCATCTACAC 1620

TGAACAAGAT TTATACAGTT ATGTCATATC TAAGCCCCGC AACAAAAGAG TACCCATTCT 1680

TCCTTTTGTT ATAGGAGCAG GAGTGCTAGG TGCACTAGGT ACTGGCATTG GCGGTATCAC 1740

AACCTCTACT CAGTTCTACT ACAAACTATC TCAAGAACTA AATGGGGACA TGGAACGGGT 1800

CGCCGACTCC CTGGTCACCT TGCAAGATCA ACTTAACTCC CTAGCAGCAG TAGTCCTTCA 1860

AAATCGAAGA GCTTTAGACT TGCTAACCGC TGAAAGAGGG GGAACCTGTT TATTTTTAGG 1920

GGAAGAATGC TGTTATTATG TTAATCAA 1948

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1136 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCATGGCCAT CTACACTGAA CAAGATTTAT ACAGTTATGT CATATCTAAG CCCCGCAACA 60

AAAGAGTACC CATTCTTCCT TTTGTTATAG GAGCAGGAGT GCTAGGTGCA CTAGGTACTG 120

GCATTGGCGG TATCACAACC TCTACTCAGT TCTACTACAA ACTATCTCAA GAACTAAATG 180

GGGACATGGA ACGGGTCGCC GACTCCCTGG TCACCTTGCA AGATCAACTT AACTCCCTAG 240

CAGCAGTAGT CCTTCAAAAT CGAAGAGCTT TAGACTCGCT AACCGCTGAA AGAGGGGGAA 300

CCTGTTTATT TTTAGGGGAA GAATGCTGTT ATTATGTTAA TCAATCCGGA ATCGTCACTG 360

AGAAAGTTAA AGAAATTCGA GATCGAATAC AACGTAGAGC AGAAGAGCTT CGAAACACTG 420

GACCCTGGGG CCTCCTCAGC CAATGGATGC CCTGGATTCT CCCCTTCTTA GGACCTCTAG 480

CAGCTATAAT ATTGCTACTC CTCTTTGGAC CCTGTATCTT TAACCTCCTT GTTAACTTTG 540

TCTCTTCCAG AATCGAAGCT GTAAAACTAC AAATGGAGCC CAAGATGCAG TCCAAGACTA 600

AGATCTACCG CAGACCCCTG GACCGGCCTG CTAGCCCACG ATCTGATGTT AATGACATCA 660

AAGGCACCCC TCCTGAGGAA ATCTCAGCTG CACAACCTCT ACTACGCCCC AATTCAGCAG 720

GAAGCAGTTA GAGCGGTCGT CGGCCAACCT CCCCAACAGC ACTTAGGTTT TCCTGTTGAG 780

ATGGGGGACT GAGAGACAGG ACTAGCTGGA TTTCCTAGGC TGACTAAGAA TCCCTAAGCC 840

TAGCTGGGAA GGTGACCACA TCCACCTTTA AACACGGGGC TTGCAACTTA GTTCACACCT 900

GACCAATCAG AGAGCTCACT AAAATGCTAA TTAGGCAAAG ACAGGAGGTA AAGAAATAGC 960

CAATCATCTA TTGCATGAGA GCACAGCAGG AGGGACAATG ATCGGGATAT AAACCCAAGT 1020

CTTCGAGCCG GCAACGGCAA CCCCCTTTGG GTCCCCTCCC TTTGTATGGG AGCTCTGTTT 1080

TCATGCTATT TCACTCTATT AAATCTTGCA GCTGCGAAAA AAAAAAAAAA AAAAAA 1136

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2782 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATGGGAGCTG TTTTCATGCT ATTTCACTCT ATTAAATCTT GCAACTGCAC TCTTCTGGTC 60

CATGTTTCTT ACGGCTCGAG CTGAGCTTTT GCTCACCGTC CACCACTGCT GTTTGCCACC 120

ACCGCAGACC TGCCGCTGAC TCCCATCCCT CTGGATCCTG CAGGGTGTCC GCTGTGCTCC 180

TGATCCAGCG AAGCGCCCAT TGCCGCTCCC AATTGGGCTA AAGGCTTGCC ATTGTTCCTG 240

CACGGCTAAG TGCCTGGGTT TGTTCTAATT GAGCTGAACA CTAGTCACTG GGTTCCATGG 300

TTCTCTTCTG TGACCCACGG CTTCTAATAG AACTATAACA CTTACCACAT GGCCCAAGAT 360

TCCATTCCTT GGAATCCGTG AGGCCAACGA ACTCCAGGTC AGAGAATACG AAGCTTGCCA 420

CCATCTTGGA AGCGGCCTGC TACCATCTTG GAAGTGGTTC ACCACCATCT TGGGAGCTCT 480

GTGAGCAAGG ACCCCCCGGT GACATTTTGG CGACCACCAA CGGACATCCC AAGTGATACA 540

TCCTGGGAAG GACCCTACCC AGTCATTTTA TCTACCCCAA CTGCGGTTAA AGTGGCTGGA 600

GTGGAGTCTT GGATACATCA CACTTGAGTC AAATCCTGGA TACTGCCAAA GGAACCTGAA 660

AATCCAGGAG ACAACGCTAG CTATTCCTGT GAACCTCTAG AGGATTTGCG CCTGCTCTTC 720

AAACAACAAC CAGGAGGAAA GTAACTAAAA TCATAAATCC CCATGGGCCT CCCTTATCAT 780

ATTTTTCTCT GTAGTGTTCT TTCACCCTGT TTCACTCTCA CTGCACCCCC TCCATGCCGC 840

TGTATGACCA GTAGCTCCCC TCACCCAGAG TTTCTATGGA GAATGCAGCG TCCCGGAAAT 900

ATTGATGCCC CATCGTATAG GAGTCTTTCT AAGGGAACCC CCACCTTCAC TGCCCACACC 960

CATATGCCCC GCAACTGCTA TCACTCTGCC ACTCTTTGCA TGCATGCAAA TACTCATTAT 1020

TGGACAGGAA AAATGATTAA TCCTAGTTGT CCTGGAGGAC TTGGAGTCAC TGTCTGTTGG 1080

ACTTACTTCA CCCAAACTGG TATGTCTGAT GGGGGTGGAG TTCAAGATCA GGCAAGAGAA 1140

AAACATGTAA AAGAAGTAAT CTCCCAACTC ACCGGGGTAC ATGGCACCTC TAGCCCCTAC 1200

AAAGGACTAG ATCTCTCAAA ACTACATGAA ACCCTCCGTA CCCATACTCG CCTGGTAAGC 1260

CTATTTAATA CCACCCTCAC TGGGCTCCAT GAGGTCTCGG CCCAAAACCC TACTAACTGT 1320

TGGATATGCC TCCCCCTGAA CTTCAGGCCA TATGTTTCAA TCCCTGTACC TGAACAATGG 1380

AACAACTTCA GCACAGAAAT AAACACCACT TCCGTTTTAG TAGGACCTCT TGTTTCCAAT 1440

GTGGAAATAA CCCATACCTC AAACCTCACC TGTGTAAAAT TTAGCAATAC TACATACACA 1500

ACCAACTCCC AATGCATCAG GTGGGTAACT CCTCCCACAC AAATAGTCTG CCTACCCTCA 1560

GGAATATTTT TTGTCTGTGG TACCTCAGCC TATCGTTGTT TGAATGGCTC TTCAGAATCT 1620

ATGTGCTTCC TCTCATTCTT AGTGCCCCCT ATGACCATCT ACACTGAACA AGATTTATAC 1680

AGTTATGTCA TATCTAAGCC CCGCAACAAA AGAGTACCCA TTCTTCCTTT TGTTATAGGA 1740

GCAGGAGTGC TAGGTGCACT AGGTACTGGC ATTGGCGGTA TCACAACCTC TACTCAGTTC 1800

TACTACAAAC TATCTCAAGA ACTAAATGGG GACATGGAAC GGGTCGCCGA CTCCCTGGTC 1860

ACCTTGCAAG ATCAACTTAA CTCCCTAGCA GCAGTAGTCC TTCGAAATCG AAGAGCTTTA 1920

GACTTGCTAA CCGCTGAGAG AGGGGGAACC TGTTTATTTT TAGGGGAAGA ATGCTGTTAT 1980

TATGTTAATC AATCCGGAAT CGTCACTGAG AAAGTTGAAG AAATTCCAGA TCGAATACAA 2040

CGTATAGCAG AGGAGCTTCG AAACACTGGA CCCTGGGGCC TCCTCAGCCG ATGGATGCCC 2100

TGGATTCTCC CCTTCTTAGG ACCTCTAGCA GCTATAATAT TGCTACTCCT CTTTGGACCC 2160

TGTATCTTTG ACCTCCTTGT TAACTTTGTC TCTTCCAGAA TCGAAGCTGT GAAACTACAA 2220

ATGGAGCCCA AGATGCAGTC CAAGACTAAG ATCTACCGCA GACCCCTGGA CCGGCCTGCT 2280

AGCCCACGAT CTGATGTTAA TGACATCAAA GGCACCCCTC CTGAGGAAAT CTCAGCTGCA 2340

CAACCTCTAC TACGCCCCAA TTCAGCAGGA AGCAGTTAGA GCGGTGGTCG GCCAACCTCC 2400

CCAACAGCAC TTAGGTTTTC CTGTTGAGAT GGGGGACTGA GAGACAGGAC TAGCTGGATT 2460

TCCTAGGCTG ACTAAGAATC CTTAAGCCTA GGTGGGAAGG TGACCACATC CACCTTTAAA 2520

CACGGGGCTT GCAACTTAGC TCACACCTGA CCAATCAGAG AGCTCACTAA AATGCTAATT 2580

AGGCAAAGAC AGGAGGTAAA GAAATAGCCA ATCATTTATT GCCTGAGAGC ACAGCAGGAG 2640

GGACAATGAT CGGGATATAA ACCCAAGTTT TCGAGCCGGC AACGGCAACC CCCTTTGGGT 2700

CCCCTCCCTT TGTATGGGAG CTCTGTTTTC ATGCTATTTC ACTCTATTAA ATCTTGCAAC 2760

TGCAAAAAAA AAAAAAAAAA AA 2782

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 666 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TGTCCGCTGT GCTCCTGATC CAGCGAGGCG CCCATTGCCG CTCCCAATTG GGCTAAAGGC 60

TTGCCATTGT TCCTGCACGG CTAAGTGCCT GGGTTTGTTC TAATTGAGCT GAACACTANT 120

CACTGGGTTC CATGGTTCTC TTCTGTGACC CACGGCTTCT AATATAACTA TAACACTTAC 180

CACATGGCCC AAGATTCCAT TCCTTGGAAT CCGTGAGGCC AAGAACTCCA GGTCAGAGAA 240

TACGAGGCTT GCCACCATCT TGGAAGCGGC CTGCTACCAT CTTGGAAGTG GTTCACCACC 300

ATCTTGGGAG CTCTGTGAGC AAGGACCCCC CGGTAACATT TTGGCAACCA CGAACGGACA 360

TCCAAAGTGA ATCGAAGCTG TAAAACTACA AATGGAGCCC AAGATGCAGT CCAAGACTAA 420

GATCTACCGC AGACCCCTGG ACCGGCCTGC TAGCCCACGA TCTGATGTTA ATGACATCAA 480

AGGCACCCCT CCTGAGGAAA TCTCAGCTGC ACAACCTCTA CTACGCCCCA ATTCAGCAGG 540

AAGCAGTTAG AGCGGTCGTC GGCCAACCTC CCCAACAGCA CTTAGGTTTT CCTGTTGAGA 600

TGGGGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCTAAGCCT 660

AGCTGG 666

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3372 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GACTTCCCAA ATACCAGAGG AAGCAGAGTG GTTTACAGTC CTGGACCTTC AGGATGCCTT 60

CTTCTGCATC CCTGTACATC CTGACTCTCA ATTCTTGTTT GCCTTTGAAG ATACTTCAAA 120

CCCAGCATCT CAACTCACCT GGACTATTTT ACCCCAAGGG TTCAGGGATA GTCCCCATCT 180

ATTTGGCCAG GCATTAGCCC AAGACTTGAG CCAATCCTCA TACCTGGACA CTTGTCCTTC 240

GGTAGGTGGA TGATTTACTT TTGGCCGCCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC 300

AAGCGCTCTT CAATTTCCTC GCTACCTGTG GCTACATGGT TTCCAAACCA AAGGCTCAAC 360

TCTGCTCACA GCAGGTTACT TAGGGCTAAA ATTATCCAAA GGCACCAGGG CCCTCAGTGA 420

GGAACACATC CAGCCTATAC TGGCTTATCC TCATCCCAAA ACCCTAAAGC AACTAAGGGG 480

ATTCCTTGGC GTAATAGGTT TCTGCCGAAA ATGGATTCCC AGGTATGGCG AAATAGCCAG 540

GTCATTAAAT ACACTAATTA AGGAAACTCA GAAAGCCAAT ACCCATTTAG TAAGATGGAC 600

AACTGAAGTA GAAGTGGCTT TCCAGGCCCT AACCCAAGCC CCAGTGTTAA GTTTGCCAAC 660

AGGGCAAGAC TTTTGTTCAT ATGTCACAGA AAAAACAGGA ATAGCTCTAG GAGTCCTTAC 720

ACAGATCCGA GGGATGAGCT TGCAACCTGT GGCACACCTG ACTAAGGAAA TTGATGTAGT 780

GGCAAAGGGT TGACCTCATT GTTTACGGGT AGTGGTGGCA GTAGCAGTCT TAGTATCTGA 840

AGCAGTTAAA ATAATACAGG GAAGAGATCT TACTGTGTGG ACATCTCATG ATGTGAATGG 900

CATACTCACT GCTAAAGGAG ACTTGTGGCT GTCAGACAAC TGTTTACTTA AATGTCAGGC 960

TCTATTACTT GAAGGGCCAG TGCTGCGACT GTGCACTTGT GCAACTCTTA ACCCAGCCAC 1020

ATTTCTTCCA GACAATGAAG AAAAGATAAA ACATAACTGT CAACAAGTAA TTTCTCAAAC 1080

CTATGCCACT CGAGGGGACC TTTTAGAGGT TCCTTTGACT GATCCCGACC TCAACTTGTA 1140

TACTGATGGA AGTTCCTTTG TAGAAAAAGG ACTTCGAAAA GTGGGGTATG CAGTGGTCAG 1200

TGATAATGGA ATACTTGAAA GTAATCCCCT CACTCCAGGA ACTAGTGCTC AGCTAGCAGA 1260

ACTAATAGCC CTCACTTGGG CACTAGAATT AGGAGAAGAA AAAAGGGCAA ATATAATACA 1320

GACTCTAAAT ATGCTTACCT AGTCCTCCAT GCCCATGCAG CAATATGGAA AGAAAGGGAA 1380

TTCCTAACTT CTGAGAGAAC ACCTATCAAA CATCAGGAAG CCATTAGGAA ATTATTATTG 1440

GCTGTACAGA AACCTAGAGA GGTGGCAGTC TTACACTGCC GGGGTCATCA CAAAGGAAAG 1500

GAAAGGGAAA TACAAGAGAA CTGCCAAGCA TATATTGAAG CCAAAAGAGC TGCAAGGCAG 1560

GACCCTCCAT TAGAAATGCT TATTAAACTT CCCTTAGTAT AGGGTAATCC CTTCCGGGAA 1620

ACCAAGCCCC AGTACTCAGC AGGAGAAACA GAATGGGGAA CCTCACGAGG CAGTTTTCTC 1680

CCCTCGGGAC GGTTAGCCAC TGAAGAAGGG AAAATACTTT TGCCTGCAAC TATCCAATGG 1740

AAATTACTTA AAACCCTTCA TCAAACCTTT CACTTAGGCA TCGATAGCAC CCATCAGATG 1800

GCCAAATCAT TATTTACTGG ACCAGGCCTT TTCAAAACTA TCAAGCAGAT AGTCAGGGCC 1860

TGTGAAGTGT GCCAGAGAAA TAATCCCCTG CCTTATCGCC AAGCTCCTTC AGGAGAACAA 1920

AGAACAGGCC ATTACCCTGG AGAAGACTGG CAACTGATTT TACCCACAAG CCCAAACCTC 1980

AGGGATTTCA GTATCTACTA GTCTGGGTAG ATACTTTCAC GGGTTGGGCA GAGGCCTTCC 2040

CCTGTAGGAC AGAAAAGGCC CAAGAGGTAA TAAAGGCACT AGTTCATGAA ATAATTCCCA 2100

GATTCGGACT TCCCCGAGGC TTACAGAGTG ACAATAGCCC TGCTTTCCAG GCCACAGTAA 2160

CCCAGGGAGT ATCCCAGGCG TTAGGTATAC GATATCACTT ACACTGCGCC TGAAGGCCAC 2220

AGTCCTCAGG GAAGGTCGAG AAAATGAATG AAACACTCAA AGGACATCTA AAAAAGCAAA 2280

CCCAGGAAAC CCACCTCACA TGGCCTGTTC TGTTGCCTAT AGCCTTAAAA AGAATCTGCA 2340

ACTTTCCCCA AAAAGCAGGA CTTAGCCCAT ACGAAATGCT GTATGGAAGG CCCTTCATAA 2400

CCAATGACCT TGTGCTTGAC CCAAGACAGC CAACTTAGTT GCAGACATCA CCTCCTTAGC 2460

CAAATATCAA CAAGTTCTTA AAACATTACA AGGAACCTAT CCCTGAGAAG AGGAAAAGAA 2520

TATTCCACCC AAGTGACATG GTATTAGTCA AGTCCCTTCC CTCTAATTCC CCATCCCTAG 2580

ATACATCCTG GGAAGGACCC TACCCAGTCA TTTTATCTAC CCCAACTGCG GTTAAAGTGG 2640

CTGGAGTGGA GTCTTGGATA CATCACACTT GAGTCAAATC CTGGATACTG CCAAAGGAAC 2700

CTGAAAATCC AGGAGACAAC GCTAGCTATT CCTGTGAACC TCTAGAGGAT TTGCGCCTGC 2760

TCTTCAAACA ACAACCAGGA GGAAAAATCG AAGCTGTAAA ACTACAAATG GAGCCCAAGA 2820

TGCAGTCCAA GACTAAGATC TACCGCAGAC CCCTGGACCG GCCTGTTAGC CCACGATCTG 2880

ATGTTAATGA CATCAAAGGC ACCCCTCCTG AGGAAATCTC AGCTGCACAA CCTCTACTAC 2940

GCCCCAATTC AGCAGGAAGC AGTTAGAGCG GTCGTCGGCC AACCTCCCCA ACAGCACTTA 3000

GGTTTTCCTG TTGAGATGGG GGACTGAGAG ACAGGACTAG CTGGATTTCC TAGGCTGATT 3060

AAGAATCCCT AAGCCTAGCT GGGAAGGTGA CCACATCCAC CTTTAAACAC GGGGCTTGCA 3120

ACTTAGCTCA CACCTGACCA ATCAGAGAGC TCACTAAAAT GCTAATTAGG CAAAGACAGG 3180

AGGTAAAGAA ATAGCCAATC ATTTATTGCC TGAGAGCACA GCAGGAGGGA CAATGATCGG 3240

GATATAAACC CAAGTTTTCG AGCCGGCAAC GGCAACCCCC TTTGGGTCCC CTCCCTTTGT 3300

ATGGGAGCTC TGTTTTCATG CTATTTCACT CTATTAAATC TTGCAACTGC AAAAAAAAAA 3360

AAAAAAAAAA AA 3372

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2372 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

ACTGCACTCT TCTGGTCCAT GTTTCTTACG GCTCGAGCTG AGCTTTTGCT CACCGTCCAC 60

CACTGCTGTT TGCCACCACC GCAGACCTGC CGCTGACTCC CATCCCTCTG GATCCTGCAG 120

GGTGTCCGCT GTGCTCCTGA TCCAGCGAGG CGCCCATTGC CGCTCCCAAT TGGGCTAAAG 180

GCTTGCCATT GTTCCTGCAC GGCTAAGTGC CTGGGTTTGT TCTAATTGAG CTGAACACTA 240

ATCACTGGGT TCCATGGTTC TCTTCTGTGA CCCACGGCTT CTAATAGAAC TATAACACTT 300

ACCACATGGC CCAAGATTCC ATTCCTTGGA ATCCGTGAGG CCAAGAACTC CAGGTCAGAG 360

AATACGAGGC TTGCCACCAT CTTGGAAGCG GCCTGCTACC GTCTTGGAAG TGGTTCACCA 420

CCATCTTGGG AGCTCTGTGA GCAAGGACCC CCCGGTAACA TTTTGGCAAC CAACGACGGA 480

CATCCAAAGT GATGGGAAAC GTTCCCCGCA AGACAAAAAC GCCCCTAAGA CGTATTCTGG 540

AGAATTGGGA CCAATTTGAC CCTCAGACAC TAAGAAAGAA ACGACTTATA TTCTTCTGCA 600

GTGCCGCCTG GCACTCCTGA GGGAAGTATA AATTATAACA CCATCTTACA GCTAGACCTC 660

TTTTGTAGAA AAGGCAAATG GAGTGAAGTG CCATAAGTAC AAACTTTCTT TTCATTAAGA 720

GACAACTCAC AATTATGTAA AAAGTGTGAT TTATGCCCTA CAGGAAGCCT TCAGAGTCTA 780

CCTCCCTATC CCAGCATCCC CGACTCCTTC CCCAACTAAT AAGGACCCCC CTTCAACCCA 840

AATGGTCCAA AAGGAGATAG ACAAAAGGGT AAACAGTGAA CCAAAGAGTG CCAATATTCC 900

CCAATTATGA CCCCTCCAAG CAGTGGGAGG AAGAGAATTC GGCCCAGCCA GAGTGCATGT 960

GCCTTTTTCT CTCCCAGACT TAAAGCAAAT AAAAACAGAC TTAGGTAAAT TCTCAGATAA 1020

CCCTGATGGC TATATTGATG TTTTACAAGG GTTAGGACAA TTCTTTGATC TGACATGGAG 1080

AGATATAATG TCACTGCTAA ATCAGACACT AACCCCAAAT GAGAGAAGTG CCACCATAAC 1140

TGCAGCCTGA GGGTTTGGCG TCTCTGGTAT CTCAGTCAGG TCAATGGATA NGGATGACAA 1200

CAGAAGGAAA GANAATGATT CCCCACAGGC CAGCAGGCAG TTCCCAGTCT AGACCCTCAT 1260

TGGGACACAG AATCAGAACA TGGAGATTGG TGCTGCAGAC ATTTGCTAAC TTGTGTGCTA 1320

GAAGGACTAA GGAAAACTAG GAAGAAGTCT ATGAATTACT CAATGATGTC CACCATAACA 1380

CAGGGAAGGG AAGAAAATCC TACTGCCTTT CTGGAGAGAC TAAGGGAGGC ATTGAGGAAG 1440

CGTGCCTCTC TGTCACCTGA CTCTTCTGAA GGCCAACTAA TCTTAAAGCG TAAGTTTATC 1500

ACTCAGTCAG CTGCAGACAT TAGAAAAAAC TTCAAAAGTC TGCCGTAGGC CCGGAGCAAA 1560

ACTTAGAAAC CCTATTGAAC TTGGCAACCT CGGTTTTTTA TAATAGAGAT CAGGAGGAGC 1620

AGGCGGAACA GGACAAACGG GATTAAAAAA AAGGCCACCG CTTTAGTCAT GACCCTCAGG 1680

CAAGTGGACT TTGGAGGCTC TGGAAAAGGG AAAAGCTGGG CAAATTGAAT GCCTAATAGG 1740

GCTTGCTTCC AGTGCGGTCT ACAAGGACAC TTTAAAAAAG ATTGTCCAAG TAGAAGTAAG 1800

CCGCCCCTTC GTCCATGCCC CTTATTTCAA GGGAATCACT GGAAGGCCCA CTGCCCCAGG 1860

GGACAAAGGT CTTTTGAGTC AGAAGCCACT AACCAGATGA TCCAGCAGCA GGACTGAGGG 1920

TGCCTGGGGC AAGCGCCATC CCATGCCATC ACCCTCACAG AGCCCTGGGT ATGCTTGACC 1980

ATTGAGGGCC AGGAAGGTTG TCTCCTGGAC ACTGGTGCGG TCTTCTTAGT CTTACTCTTC 2040

TGTCCCGGAC AACTGTCCTC CAGATCTGTC ACTATTCTGA GGGGGTCCNT AAGACGGGCA 2100

GTCACTAGAT ACTTTTTCCC AGCCACTAAG TTATGAACTG GGGAGCTTTA TTCTTTTCAC 2160

ATGCTTTTCT AATTATGCTT GAAAGCCCCA CTACCTTGTT AGGGAGAGAC ATTCTAGCAA 2220

AAGCAGGGGC CATTATACAC CTGAACATAG GAGAAGGAAC ACCCGTTTGT TGTNCCCCTG 2280

CTTGAGGAAG GAATTAATCC TGAAGTCTGG GCAACAGAAG GACAATATGG ACGAGCCAAA 2340

GAATGCCCGT CCTGTTCAAG TTAAACTAAA GG 2372

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7582 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CAACAATCGG GATATAAACC CAGGCATTCG AGCTGGCAAC AGCAGCCCCC CTTTGGGTCC 60
CTTCCCTTTG TATGGGAGCT GTTTTCATGC TATTTCACTC TATTAAATCT TGCAACTGCA 120
CTCTTCTGGT CCATGTTTCT TACGGCTCGA GCTGAGCTTT TGCTCACCGT CCACCACTGC 180
TGTTTGCCAC CACCGCANAC CTGCCGCTGA CTCCCATCCC TCTGGATCCT GCAGGGTGTC 240
CGCTGTGCTC CTGATCCAGC GARGCGCCCA TTGCCGCTCC CAATTGGGCT AAAGGCTTGC 300
CATTGTNCCT GCACGGCTAA GTGCCTGGGT TTGTTCTAAT TGAGCTGAAC ACTANTCACT 360
GGGTTCCATG GTTCTCTTCT GTGACCCACG GCTTCTAATA KAACTATAAC ACTTACCACA 420
TGGCCCAAGA TTCCATTCCT TGGAATCCGT GAGGSCAACG AACTCCAGGT CAGAGAATAC 480
GARGCTTGCC ACCATCTTGG AAGCGGCCTG CTACCRTCTT GGAAGTGGTT CACCACCATC 540
TTGGGAGCTC TGTGAGCAAG GACCCCCCGG TRACATTTTG GCRACCAMSR ACGGACATCC 600
MAAGTGATGG GAAACGTTCC CCGCAAGACA AAAACGCCCC TAAGACGTAT TCTGGARAAT 660
TGGGAMCAAT TTGACCCTCA GACACTAAGA AAGAAACGAC TTATATTCTT CTGCAGTGCC 720
GCCTGGCACT CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACYTCTTTTG 780
TAGAAAAGGC AAATGGAGTG AAGTGCCATA AGTACAAACT TTCTTTTCAT TAAGAGACAA 840
CTCACAATTA TGTAAAAAGT GTGATTTATG CCCTACAGGA AGCCTTCAGA GTCTACCTCC 900
CTATCCCAGC ATCCCCGACT CCTTCCCCAM YTAATAAGGA CCCCCCTTCA ACCCAAATGG 960
TCCAAAAGGA GATAGACAAA AGGGTAAACA GTGAACCAAA GAGTGCCAAT ATTCCCCAAT 1020
TATGACCCCT CCCAAGCAGT GGGAGGAAGA GAATTCGGCC CAGCCAGAGT GCATGTGCYT 1080
TTTYYTCTCC CAGACTTAAA GCAAATAAAA ACAGACTTAG GTAAATTCTC AGATAAYCCT 1140
GATGGCTATA TTGRTGTTTT ACAAGGGTTA GGACAATTCT TTGATCTGAC ATGGAGAGAT 1200
ATATATGTCA CTGCTAAATC AGACACTAAC CCCAAATGAG AGAAGTGCCA CCATAACTGC 1260
AGCCTGAGRG TTTGGCGATC TCTGGTATCT CAGTCAGGTC AATGGATANG GATGACAACA 1320
GAAGGAAAGA NAATGATTCC CCACAGGCCA GCARGCAGTT CCCAGTCTAS ACCCTCATTG 1380
GGGACACAGA AATCAGTAAC ATGGGAGATT GGTGCTGCAG ACATTTGCTA ACTTGTGTGC 1440
TASAAGGACT AAGGAAAACT ASGAAGAAAR TCTAYGAATT ACTCAATGAT GTCCACCATA 1500
ACACAGGGGA AGGGAAGAAA ATCCTACTGC CTTTCTGGAG AGACTAAGGG AGGCATTGAG 1560
GAAGCGTGCC TCTCTGTCAC CTGACTCTTC TGAAGGCCAA CTAATCTTAA AGCGTAAGTT 1620
TATCACTCAG TCAGCTGCAG ACATTAGAAA AAACTTCAAA AGTCTGCCGT AGGCCCGGAG 1680
CAAAACTTAG AAACCCTATT GAACTTGGCA ACYTCGGTTT TTTATAATAG AGATCAGGAG 1740
GAGCAGGCGG AACAGGACAA ACGGGATTAA AAAAAAGGCC ACCGCTTTAG TCATGACCCT 1800
CAGGCAAGTG GACTTTGGAG GCTCTGGAAA AGGGAAAAGC TGGGCAAATT GAATGCCTAA 1860
TAGGGCTTGC TTCCAGTGCG GTCTACAAGG ACACTTTAAA AAAGATTGTC CAAGTAGAAG 1920
TAAGCCGCCC CTTCGTCCAT GCCCCTTATT TCAAGGGAAT CACTGGAAGG CCCACTGCCC 1980
CAGGGGACAA AGGTCTTTTG AGTCAGAAGC CACTAACCAG ATGATCCAGC AGCAGGACTG 2040
AGGGTGCCTG GGGCAAGCGC CATCCCATGC CATCACCCTC ACAGAGCCCT GGGTATGCTT 2100

GACCATTGAG GGCCAGGAAG GTTGTCTCCT GGACACTGGT GCGGTCTTCT TAGTCTTACT 2160
CTTCTGTCCC GGACAACTGT CCTCCAGATC TGTCACTATT CTGAGGGGGT CCNTAAGACG 2220
GGCAGTCACT AGATACTTTY TCCCAGCCAC TAAGTTATGA ACTGGGGAGC TTTATTCTTT 2280
TCACATGCTT TTCTAATTAT GCTTGAAAGC CCCACTACCT TGTTAGGGAG AGACATTCTA 2340
GCAAAAGCAG GGGCCATTAT ACACCTGAAC ATAGGAGAAG GAACACCCGT TTGTTGTNCC 2400
CCTGCTTGAG GAAGGAATTA ATCCTGAAGT CTGGGCAACA GAAGGACAAT ATGGACGAGC 2460
CAAAGAATGC CCGTCCTGTT CAAGTTAAAC TAAAGGATTC CACTTCCTTT CCCTACCAAA 2520
GGCAGTACCC CCTCAGACCC AAGGCCCAAC AAGGATTCCA AAAGATTGTT AAGGACTTAA 2580
AAGCCCAAGG CTTAGTAAAA CCATGCATAA CTCCCTGCAG TAATTCCGTA GTGGATTGAG 2640
GAGGCACAGA AACCCAGTGG ACAGTGGAGG GTTAGTGCAA GATCTCAGGA TTATCAATGG 2700
AGGCCGTTGT CCTTTTATAC CCAGCTGTAC CTAGCCCTTA TACTGTGMYT TCCCAAATAC 2760
CAGAGGAAGC AGAGTGGTTT ACASTCCTGG ACCTTMAGGA TGCCTTCTTC TGCATCCCTG 2820
TACATCCTGA CTCTCAATTC TTGTTTGCCT TTGAAGATAC TTCAAACCCA RCATCTCAAC 2880
TCACCTGGAC TRTTTTACCC CAAGGGTTCA GGGATAGYCC CCATCTATTT GGCCAGGCAT 2940
TAGCCCAAGA CTTGAGYCAR TYMTCATACC TGGACACTCT TGTCCTTCRG TAKGTGGATG 3000
ATTTACTTTT RGCYGCCYRT TCAGAAACCT TGTGCCATCA AGCCACCCAA GCRCTCTTMA 3060
ATTTCCTCGC YACCTGTGGC TACAWGGTTT CCAAACSARA RGCTCARCTC TGCTCACAGC 3120
AGGTTAAATA CTTAGGRCTA ARATTATCCA AAGGCACCAR GGCCCTCAGT GAGGAAYRYA 3180
TCCAGCCTAT ACTGGCTTAT CCTCATCYCA AAACCCTAAA GCAACTAAGR GRRTTCCTTG 3240
GCRTAAYAGG YTTCTGCCGA AWATGGATTC CCCAGGTWTG GCRAAATAGC CAGGYCATTA 3300
WATACASTAA TTAAGGAAAC TCAGAAAGCC AATACCCATT TARTAAGATG GAYAMCTGAA 3360
GYMRAAGTGG CTTTCCAGGC CCCTAAAGAA GGCCTTAAAC CCAAGYCCCA GTGTTAAGYT 3420
TGCCAACRGG GCAAGACTTT TSTTYATAYR TCACAGAAAA AAACAGRAAY AGCTCTRGGA 3480
GTCCTTACAC AGRTCCRAGG GAYGAGCTTG CAACCYRTGG CRYACCTGAS TAAGGAAAYT 3540
GATGTAGTGG CAAAGGGTTG RCYTCATTGT TTAYGGGTAG TGGTGGCAGT AGCAGTYKTA 3600
GTATCTGAAG CAGTTAAAAT AATACAGGGR AGAGATCTTA CTGTGTGGAC ATCTCATGAK 3660
GTGAAYRGCA TACTCACTGC TAAAGGAGAC TTGTGGCTGT CAGACAACYG TTTACTTAAA 3720
TRTCAGGCTC TATTACTTGA ARGGCCAGTG CTGCRACTGT GCACTTGTGC AACTCTTAAC 3780
CCAGYCNCAT TTCTTCCAGA CAATGAAGAA AAGATARAAY ATAACTGTCA ACAARTAATT 3840
TCTCAAACCT ATGCCACTCG AGGGGACCTT YTAGARGTTC CYTTGACTGA TCCYGACCTT 3900
CAACTTGTAT ACTGATGGAA GTTCCTTTGT AGAAAAAGGA CTTCGAAAAG YGGGGTATGC 3960
AGTGGTCAGT GATAATGGAA TAYTTGAAAG TAATCCCCTC ACTCCAGGAA CTAGTGCTYA 4020
GCTRGCAGAA CTAATAGCCY TCAYTKGGGC ACTAGAATTA GGAGAAGRAA AAAGGGYAAA 4080
TATATATACA GACTCTRART ATGCTYACCT AGTCNTCCAT GCCCATGMRG CAATATGSAR 4140
AGAAAGGGAA TTCCTAACTT CYGAGRGAAC ACCTATCAMA CATCAGGAAG CCATTAGGAR 4200
ATTATTAYTG GCWGTACAGA AACCTARAGA GGTGGMAGTC TTACACTGCY GGGGTCATCA 4260

NAAAGGAAAG RAAAGGGAAA TASAAGRGAA YTGCCAAGCA KATATTGAAG CMAAAAGAGC 4320
TGCAAGGCAG GACCCTCCAT TAGAAATGCT TATTAAACTT CCCTTAGTAT AGGGTAATCC 4380
CTTCCGGGAA ACCAAGCCCC AGTACTCAGC AGGAGAAACA GAATGGGGAA CCTCACGAGG 4440
CAGTTTTCTC CCCTCGGGAC GGTTAGCCAC TGAAGAAGGG AAAATACTTT TGCCTGCAAC 4500
TATCCAATGG AAATTACTTA AAACCCTTCA TCAAACCTTT CACTTAGGCA TCGATAGCAC 4560
CCATCARATG GCCAAATCAT TATTTACTGG ACCAGGCCTT TTCAAAACTA TCAAGCARAT 4620
AKTCAGGGCC TGTGAAKTGT GCCARARAAA TAATCCCCTG CCTYATCGCC AAGCTCCTTC 4680
AGGARAACAA ARAACAGGCC ATTACCCTGR ARAARACTGG CAACTGATTT TACCCACAAG 4740
CCCAAACCTC AGGGATTTCA GTATCTACTA GTCTGGGTAR ATACTTTCAC GGGTTGGGCA 4800
RAGGCCTTCC CCTGTAGGAC AGAAAAGGCC CAAGAGGTAA TAAAGGCACT AGTTCATGAA 4860
ATAATTCCCA GATTCGGACT TCCCCGAGGC TTACAGAGTG ACAATAGCCC TGCTTTCCAG 4920
GCCACAGTAA CCCAGGGAGT ATCCCAGGCG TTAGGTATAC GATATCACTT ACACTGCGCC 4980
TGAAGGCCAC AGTCCTCAGG GAAGGTCGAG AAAATGAATG AAAYACTCAA AGGACATCTA 5040
AAAAAGCAAA CCCAGGAAAC CCACCTCACA TGGCCTGYTC TGTTGCCTAT AGCCTTAAAA 5100
AGAATCTGCA ACTTTCCCCA AAAAGCAGGA CTTAGCCCAT ACGAAATGCT GTATGGAAGG 5160
CCCTTCATAA CCAATGACCT TGTGCTTGAC CCAAGACAGC CAACTTAGTT GCAGACATCA 5220
CCTCCTTAGC CAAATATCAA CAAGTTCTTA AAACATTACA AGGAACCTAT CCCTGAGAAG 5280
AGGGAAAAGA ACTATTCCAC CCWWGTGACA TGGTATTAGT CAAGTCCCTT CYCTCTAATT 5340
CCCCATCCCT AGATACATCC TGGGAAGGAC CCTACCCAGT CATTTTATYT ACCCCAACTG 5400
CGGTTAAAGT GGCTGGAGTG GAGTCTTGGA TACATCACAC TTGAGTCAAA TCCTGGATAC 5460
TGCCAAAGGA ACCTGAAAAT CCAGGAGACA ACGCTAGCTA TTCCTGTGAA CCTCTAGAGG 5520
ATTTGCGCCT GCTCTTCAAA CAACAACCAG GAGGAAAGTA ACTAAAATCA TAAATCCCCC 5580
ATGGSCCTCC CTTATCATAT TTTTCTCTKT ASTGTTSTTT YACCCTSTTT CACTCTCACT 5640
GCACCCCCTC CATGCCGCTG TATGACCAGT AGCTCCCCTY ACCMAGAGTT TCTATGGAGA 5700
ATGCAGCGTC CCGGAAATAT TGATGCCCCA TCGTATAGGAG TCTTTSTAAG GGAACCCCC 5760
ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATC ACTCTGCCAC TCTTTGCATG 5820
CATGCAAATA CTCATTATTG GACAGGAAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 5880
GGAGTCACTG TCTGTTGGAC TTACTTCACC CAAACTGGTA TGTCTGATGG GGGTGGAGTT 5940
CAAGATCAGG CAAGAGAAAA ACATGTAAAA GAAGTAATCT CCCAACTCAC CSGGGTACAT 6000
GGCACCTCTA GCCCCTACAA AGGACTAGAT CTCTCAAAAC TACATGAAAC CCTCCGTACC 6060
CATACTCGCC TGGTAAGCCT ATTTAATACC ACCCTCACTG GGCTCCATGA GGTCTCGGCC 6120
CAAAACCCTA CTAACTGTTG GATATGCCTC CCCCTGAACT TCARGCCATA TGTTTCAATC 6180
CCTGTACCTG AACAATGGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 6240
GGACCTCTTG TTTCCAATST GGAAATAACC CATACCTCAA ACCTCACCTG TGTAAAATTT 6300
AGCAATACTA CATACACAAC CAACTCCCAA TGCATCAGGT GGGTAACTCC TCCCACACAA 6360
ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCGTTGTTTG 6420

AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCYAT GRCCATCTAC 6480
ACTGAACAAG ATTTATACAG TTATGTCATA TCTAAGCCCC GCAACAAAAG AGTACCCATT 6540
CTTCCTTTTG TTATAGGAGC AGGAGTGCTA GGTGCACTAG GTACTGGCAT TGGCGGTATC 6600
ACAACCTCTA CTCAGTTCTA CTACAAACTA TCTCAAGAAC TAAATGGGGA CATGGAACGG 6660
GTCGCCGACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 6720
CRAAATCGAA GAGCTTTAGA CTYGCTAACC GCTGARAGAG GGGGAACCTG TTTATTTTTA 6780
GGGGAAGAAT GCTGTTATTA TGTTAATCAA TCCGGAATCG TCACTGAGAA AGTTRAAGAA 6840
ATTCSAGATC GAATACAACG TAKAGCAGAR GAGCTTCGAA ACACTGGACC CTGGGGCCTC 6900
CTCAGCCRAT GGATGCCCTG GATTCTCCCC TTCTTAGGAC CTCTAGCAGC TATAATATTG 6960
CTACTCCTCT TTGGACCCTG TATCTTTRAC CTCCTTGTTA ACTTTGTCTC TTCCAGAATC 7020
GAAGCTGTRA AACTACAAAT GGAGCCCAAG ATGCAGTCCA AGACTAAGAT CTACCGCAGA 7080
CCCCTGGACC GGCCTGYTAG CCCACGATCT GATGTTAATG ACATCAAAGG CACCCCTCCT 7140
GAGGAAATCT CAGCTGCACA ACCTCTACTA CGCCCCAATT CAGCAGGAAG CAGTTAGAGC 7200
GGTSGTCGGC CAACCTCCCC AACAGCACTT AGGTTTTCCT GTTGAGATGG GGGACTGAGA 7260
GACAGGACTA GCTGGATTTC CTAGGCTGAY TAAGAATCCY TAAGCCTAGS TGGGAAGGTG 7320
ACCACATCCA CCTTTAAACA CGGGGCTTGC AACTTAGYTC ACACCTGACC AATCAGAGAG 7380
CTCACTAAAA TGCTAATTAG GCAAAGACAG GAGGTAAAGA AATAGCCAAT CATYTATTGC 7440
MTGAGAGCAC AGCAGGAGGG ACAATGATCG GGATATAAAC CCAAGTYTTC GAGCCGGCAA 7500
CGGCAACCCC CTTTGGGTCC CCTCCCTTTG TATGGGAGCT CTGTTTTCAT GCTATTTCAC 7560
TCTATTAAAT CTTGCARCTG CR 7582

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2563 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ACTGCACTCT TCTGGTCCAT GTTTGTTACG GCTCGAGCTG AGCTTTTGCT CGCCATCCAC 60

CACTGCTGTT TGCCACCGTT GCAGACCCAC TGCTGACTTC CATCCCTCTG GATCTGGCAG 120

GGTGTCTGCT GTGCTCCTGA TCCAGCGAGG GGCCCATTGC CACTCCCAAT CGGGCTAAAG 180

GCTTGCCATT GTTCCTGCAT GGCTAAGTGC CCAGGTTCAT CCTAATTGAG CTGAACACTA 240

GTCACTGGGT TCCACAGTTC TCTTCCATGA ACCACGGCTT TTAATAGAGC TATAACACTC 300

ATCGCAAGGC CCAAGATTCC ATTCCTTGGA ATCTGTGAGG CCAAGAACCC TAGGTCAGAG 360

AACACGAGGC TTGCCACCAT CTTGGAAGCA GCCTGCCACC ATCTGGGAAG CGGCCTGCCA 420

CCATCTTGGA AGCCGCCCGC CACCATCTTG GGAGCTCTGG GAGCAAGGAC CTCCCCGCAA 480

CCCAGTAACA TTTAGCGACC ACGAAGGGAC CTCCAAAGCG GTAATATTGG ACCACTTTCA 540

CTTGCTATTC TGTCCTATCC TTCCTTAGAA TTGGAGGAAA ATACCGGACA CCTGTCGGCC 600

GGTTAAAAAC GATTAGCGTG GCCTCCGGAC TTAAGAATCA GGTGTGAGGC TATCTGGGGA 660

AGGGCTTTCT AACAACCCCC AACCRTTCTG GGTTGGGAAT GTTGGTCTGC CTGGAGCCAG 720

CTTCCACTTT CAATTTTCCT GGGGAAGCCA AGGGCCGACT AGAGGCAGAA AGCTGTTGTC 780

CCAAATTCCC GGCAGTAGCC GGTTGAGATC ATGGCGCAGC CAGAAGTCTT TACTCCACAG 840

TCACCCATGC ATGCGCCCCT ATCTTTCCTT CTGACCCATA CCTCCTGGGT CCTAACCATG 900

ACTTTCTTAA AAGGGTAGCC CCAAAATTCT CCTTACCTCT GAATCTACTT CCTCTGATCC 960

CTGCCTCCTA GGTGCTAATG GTTCAGACTT TCATTTCCTC TAGCAAGTTG TATYTCCAAA 1020

GGGATATAAG GAAGCTCTAC ACTGTATCCT TAGGCATCTA GGCTCTAAAC CCAGGGAGTC 1080

TTGTCCCTGA TGTCCCAACC GATTTAGGTA TATAGTTCTC GACATGGGCA GTTATGTGGG 1140

ACCCATTCCC CACCACCCTT GCCAGGGCCC CAAGTTTGTA AATGGCTAAG AGAGGAAAGT 1200

GAGAGAGAGA GAGACAGAGT GAGACACAGA GAGAGGGAGA GACAGAGAGA GAGACAGAGA 1260

GGAGAGAGAC ACAGAGAGGG GAGAGACACA GAGAGGAGAA GGGGGCAGAG AGACCAAGAG 1320

GGAGTCYMAG AGAGAGAGAA AGAAGAAGAA ATAGTAGAAA AAAAAGTGTG CCCTATTCCT 1380

TTAAAAGCCA GGGTAAATTT AAAAAACCTA TACTTGATAA TTGAAGGTCT TCTCCATGAC 1440

CCTGTAACAC TCTAATACTA CCTTGTTCTC AGTGTAAACA AGGGTGTTAG CCTGAAAACA 1500

CTGAGACCGC TGACACCCAT AGCTTTCCTA TAAAAAATCC TTAACCCAGT AACCCGCAGA 1560

TGGCCCGCAT GCATTCAATC TGTAGTGGCA ACTGCTTTGC TAACAAGAAT AAAGTGGAAA 1620

AGTAACTTTT AGAGGAAACC TCATTGTGAG CACACCTCAC CAGTTCAGAA TTATTCTAAG 1680

TCAAAAAAGC AAAAAGGTAG CTTACTAACT CAAAAATCTT AAAGTATGGG GTTATTTTGT 1740

TAGAAAAAGG TAATTTAACA CTAATCACTG ATAATTCCCT TAACCCAGAA GATTTCCTAA 1800

CAGGAGATTT AAATCTTAAT TACCATACAA AGGTCTGACC AGACCTAGGA GGAACTCCCT 1860

TCAGTACAGG ATGATAGATG GTTCCTCCCA GGTGAATGAA AAAAAAATCA CAATGGGTAT 1920

TCAGTAATTG ATAGGGAGAC TCTTGTGGAA GCAGAGTTAG AAAAACTGCC TAATAATTGG 1980

TCTCCCCAAA CCTGCGAGCT GTTTGCACTC AGCCAAGCCT TAAAGTACTT CTAGAATCAA 2040

AAAGATTATC TCAATCCTGA CTCAAAAGGT TACCTACACC CTCTGTGAAA CGAATTTACT 2100

TAAGAACTGT TTATGGGACT GCATCTTGAT GGGGCAGCTG GGTTGTCATG AAATACTCAG 2160

GAATGCAGCC TAGCTCTAGG ACTCACCCCT GAGCACAAAG GCAATGTTGG GCATGCTGGT 2220

AAAGGACCAC TAGAATCCAG CAGTCCGAAC CCTTTCTTTG GGTTAAGAAA GGCGGGAAAA 2280

CAGGCGCAGG ACTGCTACAT TGGTAAGCGT AACTAATCCA ATAAGCAGAG GTCCATGGGT 2340

GGTGACACAC TCTGGAAAGG AATAAGCATT AGRACCATAG AGGACGCTCT ACGACTAATG 2400

CTCGTCGGAA AATGACTAGA GGTGCTGGCA TCCCTATGTT CTTTTTTCAG ATGGGAAATG 2460

TTCCCCCTCA AGGCAAAAAC ACCCCTAAGA TGTATTCTGG ACAATTGGGA CCAATTTGAC 2520

CCTCAGACTC TAAGAAAGAA ACGACTTATA TTCTTCTGCA GTG 2563

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2585 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TCAGGGATAG CCCCCATCTA TTTGGCCAGG TATTAGCCCA AGACTTGAGC CAGTTCTCAT 60

ACTTGGACAC TCTTGTCCTT TGGTATGTGG ATGATCTACT TTTAGCCACC TGTTCAGAAA 120

CCTTGTGCCA TCAAGCCAAC CAAGTGCTCT TAAACTTCCT CGCCACCTGT GGCTACAAGG 180

TTTCCAAACC AGAGGCTCAG CTCTGCTTAC AGCAGGTTAA ATACTTAGGG CTAAAATTAT 240

CCAAAGGCAC CAGGGCCCTC AGTGAGGAAC GTATCCAGCC TATACTGGCT TATCCTCATC 300

CCAAAACCCT GAAGCAATTA AGAGGGTTCC TTGGCATAAA AGGCTGCTGT TGAATATGGA 360

TTCCCAGGTA CAATGAAATA GCCAGGCCAT TATACACACT AATTACGGGA ACTCAGAAAG 420

CCAATACCCA TTTAGTAGAA TGGACACCTG AAGCAGAAGC GGCTTTCCAG GCCCTAAAGA 480

AGGCCCTAAT CCAAGCCCCA GTGTTAAGCT TGCCAATGGA GCAAGACTTT TCTTTATATG 540

TCACAGAAAA AAAAACAGGA ATAGCTCTAG AAGTCCTTAC ACAGGTCCGA GGGACCAGCT 600

TACAACACAT GGCATACCTG AGTAAGGAAA CTGATGTAGT GGCAAAGGGT TGGACTCATT 660

GTTTACAGGT AGTGGCAGCA GTAGCAGTCT TAGCATCTGA AGCAGTTAAA ATGATACAGG 720

GAAGANATCT TACTGTGTGG ACATCTCATG ATGTGAACGG CATACTCACT GCTAAAGGAG 780

ACTGTGGCTG TCAGACAACC ATTTGCTTAA ATATCAGGCT CTATCACTTG AANGGCCAGT 840

GCTGCCACTG TGCACTTGTG CAACTCTTAA CCCACCCACA TTTCTTCCAG ACAATGAAGA 900

AAAGATAGAA CATAACTGTC AACAAGTGAT TGTTCAAACC TACACCGCTC GAAGGGACCT 960

TCTAGAGGTT CCCTTGACTG ATCCTGAGCT CAACTTCTAT ACTGATGGAA GTTCCTTTTG 1020

TAGAAAAAGG ACTTCGAAAG GCGGGTATGC AGTGGCCAGT GATAATGGAA TACTTGAAAG 1080

TAATCCCTTC ACTCCAGAAA CTAGCATTCA GCTGGCAGAA TTAATAGCCT TCACTTGGGC 1140

ATTAGAACAC AGGAGAAGGA AAAGGAGTAA ATATATATAC AGACTCCAAG TATGCTTACT 1200

TAGTCCTCCA TGCCCATGCA GCAATATAGA GAGAAAGCGA ATTCCTAACT TCTGAGGGAA 1260

CACCTATCAA ACATCAGGAA GCCATTAGGA GATTATTACT GGCTGTACAG AAACCTAGAG 1320

GTGGCAGTCT TACATGGCCG AGATCATCAG AAAGGAAAAG AAAGGGAAAT AGAAGGGAAC 1380

TGCCAAGTGG ATATTGAAGC CAAAAGAGCT GCAAGGCGGG ACCCTCCATT AGAAATGCTT 1440

ATAGAAGGAC CCCTAGTACA GGGCAATCCC CTTCAGGAAA CCAAGCCCCA ATACTCAGCA 1500

GAAGAAATGG AATGGGGAAC CTCATGAGGA CATAGTTTCC TCCCCTCAGG ATGGCTAGCC 1560

ACCAAAGAAG GAAAAATACT TTTGCCTGCA GCTAACCAAT GGAAATTACT TAAAACCCTT 1620

CACCAAACCT TTCGCTTAGG CATTGATAGC ACCCATCAGA TGGCTAAATC ATTATTTACT 1680

AGACCACACC TTTTCAAAAC TATCAAGCAG ACAGTTAGGG CCTGTGAAGT GTGCCAAAGA 1740

AATAATCCCC TGCCTTATCG CCAAACTCCT TCAGGAGAAA AAAGAACAGG CCATTACCCA 1800

GGAGAAGAGT GGCAACTAGA TTTTACCCAC ATGCCCAAAT CTCAGGGATT TCAGTATCTA 1860

CTAGTCTGGG TAGATACTTT CACTGGTTGG GCGGAGGCCT TCCCTTGTAG GACAGAACAG 1920

GCCCATGAGG TAATAAAGGC ACTAATTCAT GAAATAATTC CCAGATTTGG ATTTCCCCAA 1980

GGCTTACAGA GTGATAACGG CCCCACTTTC AAGGCTACAG TAACCCAGGG AGTATCCCAG 2040

ACATTAGACA TACAATATCA CTTACACTGA GCCCGGAGGC CACAATCCTC AGGAAAGTTG 2100

AGAAAATGAA TGAAACGCTC AAATGACATC TAAAAAAGCT AACCTAAGAA ACCCACCTCT 2160

CATGGTTTGC TCTGTTGCCT ATAGCCTTAG TAAGAATCCG AAACTCTCCC CAAAAAGCGG 2220

GACTCAGCCC ATACGAAATG CTGTATGGAC GGCCCTTCCT AACCAATGAC CTTGTGCTTG 2280

ACCTAGAGAT GGCCAACTTA GTTGCAGATA TCCCTCCTTA GCCAAATATC AACAAGTTCT 2340

TAAAACGTCA CAGGGAACCT GTCCCTGAGA GGAGGGAAAG GAATTATTCC AACCTGGTGA 2400

CATGGTATTA GTGAAGTCCC TTCCCTCCAA CTCCCCATCC CCTGGATACA TCCTGGGAAG 2460

GACCCTACTC AGTCATTTTA TCTATCCCAA CCGCGGTTAA AATGGCTGGA GTAGAATCTT 2520

GGATACATCA CATTCGAGTC AAACCCTAGA TACTGCCACA AGGAACCTGA AAATCCAGGA 2580

GACAA 2585

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2575 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GGGATAGCCC CCATCTATTT GGCCAGGCAT TAGCCCAAGA CTTGAAGCCA ATTCTCATAC 60

CTGGACACTC TTCTCCTTTG GTATGTGGAT GATTTACTTT TAGCTTCCTG TTCAGAAACC 120

TTGTGCCATC AAGCCACCCA AGCACTCTTA AATTTCCTCG CTACCTGTGG CTACAAGGTT 180

TCCAAACCAA AGACCCAGCT CTGCTCACAG CAGGTTAAAT ACTTGGGGCT AAAATTATCC 240

AAAGGCACCA GGGCCCTCAG TGAGGAACGT ATCAAGCCTA TACTGGCTTA TCCTCATCCC 300

CAAATCCTAA AGCAACTAAG AGAGTTCCTT AGCATAACAG GTTTCTGCTG AATATGGATT 360

CCCAGGTATG GCAAAATAGC CAGACCATTA TATACGCTAA TTAAGGAAAC TCAGAAAGCC 420

AATACCCATT TAGTAAGATG GATACCTGAA GCAGAAGCAG CTTTCCAGGC CCTAAAGAGG 480

GCCCTAACCC AAGCCCCAGT GTTAAGCTTG CCAACAGGGC AAGACTTTAC TTCGTATGTC 540

ACAGAAAAAA CAGGAAATAG CTCTAGGAGT CCTTACACAA GTCTGAGGGA TGAGCTTGCA 600

ACCCATGGCA TACCTGAGTA AGGAAATTGA TGTAGTGGCA AAGGGTTGGC CTCATTGTTT 660

ATGGGTAGTG GCGGCAGTAG CAGTCTTAGC ATCTGAAGCA GTTAAAATGA TACAGGGAAG 720

AGATCTTACT GTGTGGACAT CTCATGATGT GAATGGCATA CTCACTGCTA AAGGAGACTT 780

GTGGCTGTCA GACAACCATT TACTTAAATA TCAGGCTGTA TTACTTGAAG GGCCAGTGCA 840

GCAACTGCGC AGTTGTGCAG CTCTTAACCC AGCCACATTT CTTCCAGACA ATGAAGATAG 900

AACATAACTG CCAACAAGTA ATTTCTCAAA CCTAGGCCGC TCGAGGGAAC CTTTTAGAGG 960

TTCCCTTAAC TGATCCCGAC CTCAACTTGT ATACTGATGG AAGTTCCTTT GTAGAAAAAG 1020

GACTTTGAAA AGTGGGGTAT GCAGTGCTCA GTGATAATGG AATACTTGAA AATAATCCCT 1080

TCATTCCAGG AACCAGCGTT CAGCTGGCAG AATTAATAGC CCTCACTCGG GCATTAGAAT 1140

TAGGAGAAGG AAAAAGGGTA AATACACATA CAGATTCTAA GTATGTTTAC TTAGTCCTCC 1200

GTGCCCACGC AGCAATATGG AGAGAAAGGG AATGCTTAAC TTCTGAGGGA ACACCTATCA 1260

AACATCAGGA AGTTATTAGG AGATTATTAT TGGCTATACA GAAACCTAAA GAGGTGGCAG 1320

TCTTACACTG CTGGGGTGGT CAGAAAGAAA AGGAAAGGGA AATAAAAGGG AACTGCCAAG 1380

CGGATATTGA AGCCAAAAGA GCCGCAAGGC AGGACCCTCC ATTAGAAATG CTTATAGAAG 1440

GACCCCTAGT ATGGGGTAAT CCCCTCCGGG AAACCAAGCC CCAATACTTA GAAAAAGAAA 1500

TAGAATGGGG AACCTCACGA GGACATAGTT TCCTCCCCTC AGGATGGCTA GCCACCGAAG 1560

AAGGAAAAAT ACTTTTGCCT GCAGCTAACC AATGGAAATT ACTTAAAACC CTTCACCAAA 1620

CCTTTCACTT AGACATTGAT AGCACCCATC AGATGGCCAA ATCATTATTT ACTGGACCAG 1680

GCCTTTTCAA AACTATCAAG CAGCTAGTCA GGGCCTGTGA AGTGTGCCGA AGAAATAATC 1740

CCATGCCTTA TCACCAAGCT CCTTCAGGAG AACAAAGAAC AGGCCATTAC CCAGGAGAAG 1800

RVTGGCAACT AGATTTTACC CACATGCCCA AATCTCAGGG ATTTCAGTAT CTACTAGTTT 1860

GGGTAGATAC TTTCACTGGT TGGGCAGAGA CCTTCCCCTG TAAGACAGAA AAGTCCCAAG 1920

AGGTAATAAA GGCATTAGTT CATGAAATAA TTCCCAGATT CAGACTTCCC TGAGGCTTAC 1980

AGAGTGACAA TGGCCCTGCT TTCAAGGCTA CAGTAACCCA GGAGTATCCC AGGTGTTAGG 2040

TATACAATAT CACTTACACT GCGCCTGGAG GCAGTCCTCA GGGAAGGCCG AGAAACTGAA 2100

TGAAACACTC AAACGACATC TAAAAAAAGC TAACCCAGGA AAACCACCTC ACATGGCCTG 2160

CTCTGTTGCC TATAGCCTTA CTAAGAATCC AAAACTCTCC CCAAAAAGCA GGACTTAGCC 2220

CATACGAAAT GCTATATGGA TAGCCCTTCC TAACCAATGA CCTTGTGCTT GACTGAGAGA 2280

GAGCCAACTT AGTTGCAGAC ATCACCTCCT TATCCAAATA TCAACAAGTT CTTAAAACAT 2340

TACAAGGAGC CTGTCCCCGA GAAGAGGGGA AGGAACTATT CCACCCTGGT GACATGGTAT 2400

TAGTCAAGTC CCTTCCCTCT AATTCTCATT GCCTAGATAT ATCCTGGGAA GGACCCTACC 2460

CAGTCATTTT ATCTACCCCA ACCGCAGTAA AAGTGGCTGG AGTGGAGTCT TGGATACATC 2520

ACACTCGAGT CAAACCCTGG ATATTACCAA AGGAACCTGA AAATCCAGGA GACAA 2575

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 783 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
TGAGAGACAG GACTAGCTGG ATTTCCTAGG CYGACTAAGA ATCCYTAAGC CTAGSTGGGA  60
AGGTGACCAC RTCCACCTTT AAACACGGGG CTTGCAACTT AGYTCACACC TGACCAATCA 120
GAGAGCTCAC TAAAATGCTA ATTAGGCAAA GACAGGAGGT AAAGAAATAG CCAATCATYT 180
ATTGCMTGAG AGCACAGCAG GAGGGACAAY RATCGGGATA TAAACCCARG YHTTCGAGCY 240
GGCAACRGCA GMCCCCCTTT GGGTCCCYTC CCTTTGTATG GGAGCTCTGT TTTCATGCTA 300
TTTCACTCTA TTAAATCTTG CARCTGCRCT CTTCTGGTCC ATGTTTCTTA CGGCTYGAGC 360
TGAGCTTTYG CTCRCCRTCC ACCACTGCTG TTTGCCRCCA CCGCANACCY GCCGCTGACT 420
CCCATCCCTC TGGATCMTGC AGGGTGTCCG CTGTGCTCCT GATCCAGCGA RGCRCCCATT 480
GCCGCTCCCA ATYGGGCTAA AGGCTTGCCA TTGTNCCTGC AYGGCTAAGT GCCTGGGTTY 540
RTYCTAATTG AGCTGAACAC TANTCACTGG GTTCCATGGT TCTCTTCTGT GACCCACRGC 600
TTCTAATAGA RCTATAACAC TYACCRCATG GCCCAAGRTT CCATTCCTTG GAATCCRTRA 660
RGSCAACGAA CYCCASGTCA GAGAAYACGA RGCTTGCCAC CATCTTGGAA GCGGCCTGCT 720
ACCATCTTGG AAGTGGTTCA CCACCATCTT GGGAGCTCTG TGAGCAAGGA CCCCCMRGTR 780
ACA                                                               783
```

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TGTCCGCTGT GCTCCTGATC 20

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATGCACTCTG GCTGGGCCAA T 21

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

ACCATTTGAC CCTCAGACAC T 21

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AACCCTTTGC CACTACATCA ATTT 24

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TCAGGGATAG CCCCCATCTA T 21

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TTGTCTCCTG GATTTTCAGG TT 22

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGACCCTACC CAGTCATTTT 20

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleotide

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ATCAGGAGCA CAGCGGACAC 20

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GGACATCCAA AGTGATACAT CC 22

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AATGTATGGC CTGAAGTGCA G 21

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CTTCCCAGGA TGTATCACTT TG 22

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CACTGCAGAA GAATATAAGT CGTT 24

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GCTTCCAAGA TGGTGGCAAG C 21

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 678 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TCAGGGATAG CCCCCATCTA TTTGGCCAGG CATTAGCCCA AGACTTGAGC CAGTTCTCAT 60

ACCTGGATAT TCTTGTCCTT TGGTATGCGG ATGATTTACT TTTAGCCGCC CGTTCAGAAA 120

CCTTGTGCCA TCAAGCCACC CAAGTGCTCT TAAATTTCCT CGCCACCTGT GGCTACAAGG 180

TTTCCAAACC AAAGGCTCAG CTCTGCTCAC AGCAGAAGGC TATTTACCCT AAATACTTAG 240

GGCTGAAATT ATCCAAAGGC ACCAGGGCCC TCAGTGAGGA ATGTATCCAG CCTATACTGG 300

CTTATCCTTA TCCCAAAACC CTAAAACAAC TAAGAAGGTT CCTTGGCATA ATAGGCATAA 360

CAGGCATAAC AGGTTTCTGC TGAATATGGA TTCCCAAGTA CGGCAAAATA GCCAGACCAT 420

TATATACACT AATTAAGGAA ACTCAGAAAG CCAATACCCA TTTAGTAAGA TGGACACCTG 480

AAGCAGAGGC AGCTTTCCAG GCCGTAAAGA ACACCCTAAC CCAAGCCCCA GTGTTAAGCT 540

TGCCAGCGGG GCAAGACTTT TCTTTCTGTG TCACAGAAAA AATAGGAATA GCTNTAGGAG 600

TCCTTACACA GGTCCGAGGG ACCAGCTTGC AACCCATGGC ATACCTGAGT AAGGAAATTG 660

ATGTAGTGGC AAAGGGTT 678

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 536 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

```
CCAATCTCCA TGTTGTATCC CCTTCCCCAA CTAATAAGGA CCCCCCTTTC AACCCAAACA   60

GTCCAAAAGG ACATAGACAA AGGAGTAAAC AATGAACCAA AGAGTGCCAA TATTCCCTGG  120

TTATGCACCC TCCAAGCGGT GGGAGAAGAA TTCGGCCCAG CCAGAGTGCA TGTACCTTTT  180

TCTCTCTCAC ACTTGAAGCA AATTAAAATA GACCTAGGTA AATTCTCAGA TAGCCCTGAT  240

GGCTATATTG ATGTTTTACA AGGATTAGGA CAATCCTTTG ATCTGACATG GAGAGATATA  300

ATATTACTGC TAAATCAGAC GCTAACCTCA AATGAGAGAA GTGCTGCCAT AACTGGAGCC  360

CGAGAGTTTG GCAATCTCTG GTATCTCAGT CAGGTCAATG ATAGGATGAC AACGGAGGAA  420

AGAGAACGAT TCCCCACAGG GCAGCAGGCA GTTCCCAGTG TAGCTCCTCA TTGGGACACA  480

GAATCAGAAC ATGGAGATTG GTGCCGCAGA CATTTAAAGC TTTCCCCGGG TACCGA      536
```

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 591 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCATGGCCAT CTACACTGAA CAAGATTTAT ACAATCATGT CGTACCTAAG CCCCACAACA 60

AAAGAGTACC CATTCTTCCT TTTGTTATCA GAGCAGGAGT GCTAGGCAGA CTAGGTACTG 120

GCATTGGCAG TATCACAACC TCTACTCAGT TCTACTACAA ACTATCTCAA GAAATAAATG 180

GTGACATGGA ACAGGTCACT GACTCCCTGG TCACCTTGCA AGATCAACTT AACTCCCTAG 240

CAGCAGTAGT CCTTCAAAAT CGAAGAGCTT TAGACTTGCT AACCGCCAAA AGAGGGGGAA 300

CCTGTTTATT TTTAGGAGAA GAACGCTGTT ATTATGTTAA TCAATCCAGA ATTGTCACTG 360

AGAAAGTTAA AGAAATTCGA GATCGAATAC AATGTAGAGC AGAGGAGCTT CAAAACACCG 420

AACGCTGGGG CCTCCTCAGC CAATGGATGC CCTGGGTTCT CCCCTTCTTA GGACCTCTAG 480

CAGCTCTAAT ATTGTTACTC CTCTTTGGAC CCTGTATCTT TAACCTCCTT GTTAAGTTTG 540

TCTCTTCCAG AATTGAAGCT GTAAAGCTAC AGATGGTCTT ACAAATCTAG A 591

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CTAACCTGAG GATCCAGCAG CAGGACTGAG GGTGCCCGGG GCAAGTGCCA GCCCATGCCA 60

TCACCCTCAG AGCCCCGGGT ATGTTTGACC ATTGAGAGCC AGGAAGTTAA CTGTCTCCTG 120

GACACTGGCG CAGCCTTCTC AGTCTTACTT TCCTGTCCCA GACAATTGTC CTCCAGATCT 180

GTCACTATCC GAGGGGTCCT AGGACAGCCA GTCACTACAT ACTTCTCTCA GCCACTAAGT 240

TGTGACTGGG GAACTTTACT CTTTTCACAT GCTTTTCTAA TTATGCCTGA AAGCCCCACT 300

CCCTTGTTAG GGAGAGACAT TTTAGCAAAA GCAGGGGCCA TTATACACCT GAACAAGCTT 360

GAAA 364

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 538 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Met Gly Leu Pro Tyr His Ile Phe Leu Cys Ser Val Leu Ser Pro Cys
1 5 10 15

Phe Thr Leu Thr Ala Pro Pro Pro Cys Arg Cys Met Thr Ser Ser Ser
20 25 30

Pro His Pro Glu Phe Leu Trp Arg Met Gln Arg Pro Gly Asn Ile Asp
35 40 45

Ala Pro Ser Tyr Arg Ser Leu Ser Lys Gly Thr Pro Thr Phe Thr Ala
50 55 60

His Thr His Met Pro Arg Asn Cys Tyr His Ser Ala Thr Leu Cys Met
65 70 75 80

His Ala Asn Thr His Tyr Trp Thr Gly Lys Met Ile Asn Pro Ser Cys
85 90 95

Pro Gly Gly Leu Gly Val Thr Val Cys Trp Thr Tyr Phe Thr Gln Thr
100 105 110

Gly Met Ser Asp Gly Gly Gly Val Gln Asp Gln Ala Arg Glu Lys His
115 120 125

Val Lys Glu Val Ile Ser Gln Leu Thr Gly Val His Gly Thr Ser Ser
130 135 140

Pro Tyr Lys Gly Leu Asp Leu Ser Lys Leu His Glu Thr Leu Arg Thr
145 150 155 160

```
His Thr Arg Leu Val Ser Leu Phe Asn Thr Thr Leu Thr Gly Leu His
                165                 170                 175

Glu Val Ser Ala Gln Asn Pro Thr Asn Cys Trp Ile Cys Leu Pro Leu
            180                 185                 190

Asn Phe Arg Pro Tyr Val Ser Ile Pro Val Pro Glu Gln Trp Asn Asn
        195                 200                 205

Phe Ser Thr Glu Ile Asn Thr Thr Ser Val Leu Val Gly Pro Leu Val
    210                 215                 220

Ser Asn Val Glu Ile Thr His Thr Ser Asn Leu Thr Cys Val Lys Phe
225                 230                 235                 240

Ser Asn Thr Thr Tyr Thr Thr Asn Ser Gln Cys Ile Arg Trp Val Thr
                245                 250                 255

Pro Pro Thr Gln Ile Val Cys Leu Pro Ser Gly Ile Phe Phe Val Cys
            260                 265                 270

Gly Thr Ser Ala Tyr Arg Cys Leu Asn Gly Ser Ser Glu Ser Met Cys
        275                 280                 285

Phe Leu Ser Phe Leu Val Pro Pro Met Thr Ile Tyr Thr Glu Gln Asp
    290                 295                 300

Leu Tyr Ser Tyr Val Ile Ser Lys Pro Arg Asn Lys Arg Val Pro Ile
305                 310                 315                 320

Leu Pro Phe Val Ile Gly Ala Gly Val Leu Gly Ala Leu Gly Thr Gly
                325                 330                 335

Ile Gly Gly Ile Thr Thr Ser Thr Gln Phe Tyr Tyr Lys Leu Ser Gln
            340                 345                 350
```

```
Glu Leu Asn Gly Asp Met Glu Arg Val Ala Asp Ser Leu Val Thr Leu
        355                 360                 365

Gln Asp Gln Leu Asn Ser Leu Ala Ala Val Val Leu Arg Asn Arg Arg
    370                 375                 380

Ala Leu Asp Leu Leu Thr Ala Glu Arg Gly Gly Thr Cys Leu Phe Leu
385                 390                 395                 400

Gly Glu Glu Cys Cys Tyr Tyr Val Asn Gln Ser Gly Ile Val Thr Glu
                405                 410                 415

Lys Val Glu Glu Ile Pro Asp Arg Ile Gln Arg Ile Ala Glu Glu Leu
            420                 425                 430

Arg Asn Thr Gly Pro Trp Gly Leu Leu Ser Arg Trp Met Pro Trp Ile
        435                 440                 445

Leu Pro Phe Leu Gly Pro Leu Ala Ala Ile Ile Leu Leu Leu Leu Phe
    450                 455                 460

Gly Pro Cys Ile Phe Asp Leu Leu Val Asn Phe Val Ser Ser Arg Ile
465                 470                 475                 480

Glu Ala Val Lys Leu Gln Met Glu Pro Lys Met Gln Ser Lys Thr Lys
                485                 490                 495

Ile Tyr Arg Arg Pro Leu Asp Arg Pro Ala Ser Pro Arg Ser Asp Val
            500                 505                 510

Asn Asp Ile Lys Gly Thr Pro Pro Glu Glu Ile Ser Ala Ala Gln Pro
        515                 520                 525

Leu Leu Arg Pro Asn Ser Ala Gly Ser Ser
    530                 535
```

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Met Glu Pro Lys Met Gln Ser Lys Thr Lys Ile Tyr Arg Arg Pro Leu
1 5 10 15

Asp Arg Pro Ala Ser Pro Arg Ser Asp Val Asn Asp Ile Lys Gly Thr
20 25 30

Pro Pro Glu Glu Ile Ser Ala Ala Gln Pro Leu Leu Arg Pro Asn Ser
35 40 45

Ala Gly Ser Ser
50

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

Met	Leu	Met	Thr	Ser	Lys	Ala	Pro	Leu	Leu	Arg	Lys	Ser	Gln	Leu	His
1				5					10					15	

Asn	Leu	Tyr	Tyr	Ala	Pro	Ile	Gln	Gln	Glu	Ala	Val	Arg	Ala	Val	Val
			20					25					30		

Gly	Gln	Pro	Pro	Gln	Gln	His	Leu	Gly	Phe	Pro	Val	Glu	Met	Gly	Asp
		35					40					45			